[illegible]

91
11.8.70

BUREAU OF EDNL. & PSYL. RESEARCH.
LIBRARY.

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
....
Dated ....
Accs. No ....

382

JOURNAL OF CONSULTING PSYCHOLOGY

Edited by
Laurance F. Shaffer
Teachers College, Columbia University

Associate Editors

Edgar A. Doll
Bellingham, Washington

Ann Magaret Garner
University of Illinois

William A. Hunt
Northwestern University

E. Lowell Kelly
University of Michigan

Morris Krugman
Board of Education
New York, N. Y.

Bertha M. Luckey
Board of Education
Cleveland, Ohio

Fred McKinney
University of Missouri

Anne Roe
New York, N. Y.

Carl R. Rogers
University of Chicago

R. Nevitt Sanford
University of California

Edward Joseph Shoben, Jr.
Teachers College
Columbia University

J. R. Wittenborn
Yale University

Managing Editor
Lorraine Bouthilet

VOLUME 17
1953

Published Bimonthly By
The American Psychological Association, Inc.
1333 Sixteenth St., N. W.
Washington 6, D. C.

Printed At
Twenty-Three West Colorado Avenue
Colorado Springs,, Colorado

Copyright, 1953, by the American Psychological Association, Inc.

Edn
11. 8. 70
91

Contents of Volume 17

Allen, R. M., Ray, C. D., and Poole, R. C. The Levy Movement Test: Suggestions for Scoring and Relationship to Rorschach Movement Responses - - 195
Anderson, Helen J. See *Machover, S.*
Armitage, S. G. See *Daston, P. G.*
Auld, F., Jr., and Eron, L. D. The Use of Rorschach Scores to Predict whether Patients Will Continue Psychotherapy - - - - - - - - - - - - - - - 104

Barrell, R. P. Subcategories of Rorschach Human Movement Responses: A Classification System and Some Experimental Results - - - - - - - - - 254
Barron, F. Some Test Correlates of Response to Psychotherapy - - - - - - - - 235
Barron, F. An Ego-Strength Scale Which Predicts Response to Psychotherapy - - - 327
Bayard, Jean. See *Pascal, G. R.*
Beier, E. G., Izard, C. E., Smock, C. D., and Tougas, R. R. Response to the Human Face as a Standard Stimulus - - - - - - - - - - - - - 126
Beloff, H. See *Cattell, R. B.*
Berger, D. The Rorschach as a Measure of Real-Life Stress - - - - - - - - - 355
Berkowitz, M., and Levine, J. Rorschach Scoring Categories as Diagnostic "Signs" - - 110
Bijou, S. W. See *Kenny, D. T.*
Bills, R. E. Rorschach Characteristics of Persons Scoring High and Low in Acceptance of Self - 36
Bills, R. E. A Validation of Changes in Scores on the Index of Adjustment and Values as Measures of Changes in Emotionality - - - - - - - 135
Bills, R. E. A Comparison of Scores on the Index of Adjustment and Values with Behavior in Level-of-Aspiration Tasks - - - - - - 206
Botwinick, J. Wechsler-Bellevue Split-Half Subtest Reliabilities: Differences in Age and Mental Status - - - - - - - - - - - - - - 225

Calvin, A. D., and Holtzman, W. H. Adjustment and the Discrepancy between Self Concept and Inferred Self - - - - - - - - - - - - - 39
Calvin, A. D., and McConnell, J. Ellis on Personality Inventories - - - - - - 462
Canter, F. M. Personality Factors in Seizure States with Reference to the Rosenzweig Triadic Hypothesis - - - - - - - - - - - - - 429
Carlson, V. R., and Lazarus, R. S. A Repetition of Meyer Williams' Study of Intellectual Control under Stress and Associated Rorschach Factors - - - - 247
Cattell, R. B., and Beloff, H. Research Origin and Construction of the I. P. A. T. Junior Personality Quiz - - - - - - - - - - - 436
Cole, D. Communication and Rapport in Clinical Testing - - - - - - - - - 132
Cole, Mary E. See *Pascal, G. R.*
Collier, R. M. A Scale for Rating the Responses of the Psychotherapist - - - - - 321
Cook, D. M. See *Strother, G. B.*
Cowen, E. L., Wiener, M., and Hess, Judith. Generalization of Problem-Solving Rigidity 100
Crown, S. Psychological Changes following Operations on the Human Frontal Lobe - 92
Dailey, C. A. The Practical Utility of the Clinical Report - - - - - - - - - - 297

Dalton, P. G., King, G. F., and Armitage, S. G. Distortion in Paranoid Schizophrenia - 50
Dupont, H. J., Landsman, T., and Valentine, M. The Treatment of Delayed Speech by Client-Centered Therapy - - - - - - - - - - - - - - 122
Dymond, Rosalind F. An Adjustment Score for *Q* Sorts - - - - - - - - - - - - 339

Ellis, A. Recent Research with Personality Inventories - - - - - - - - - - - - 45
Epstein, S. Overinclusive Thinking in a Schizophrenic and a Control Group - - - - 384
Eron, L. D. Responses of Women to the Thematic Apperception Test - - - - - - 269
Eron, L. D. See *Auld, F., Jr.*
Estes, Betsy Worth. Influence of Socioeconomic Status on Wechsler Intelligence Scale for Children: An Exploratory Study - - - - - - - - - - - 58

Feifel, H., and Schwartz, A. D. Group Psychotherapy with Acutely Disturbed Psychotic Patients - 113
Feldman, Dorothy A. See *Pascal, G. R.*

Galanter, E. H. See *Levine, M.*
Gallagher, J. J. Clinical Judgment and the Cattell Infant Intelligence Scale - - - - 303
Gallagher, J. J. MMPI Changes Concomitant with Client-Centered Therapy - - - 334
Gallagher, J. J. Manifest Anxiety Changes Concomitant with Client-Centered Therapy - 443
Gibby, R. G., Miller, D. R., and Walker, E. L. The Examiner's Influence on the Rorschach Protocol - - - - - - - - - - - - - - - - 425
Gibby, R. G., and Stotsky, B. A. The Relation of Rorschach Free Association to Inquiry 359
Gibby, R. G., Stotsky, B. A., Miller, D. R., and Hiler, E. W. Prediction of Duration of Therapy from the Rorschach Test - - - - - - - - 348
Gough, H. G. A Nonintellectual Intelligence Test - - - - - - - - - - - - - - - 242
Graham, E. E., and Shapiro, Esther. Use of the Performance Scale of the Wechsler Intelligence Scale for Children with the Deaf Child - - - - - - - - 396
Granick, S., and Smith, L. J. Sex Sequence in the Draw-a-Person Test and Its Relation to the MMPI Masculinity-Femininity Scale - - - - - - 71
Grauer, D. Prognosis in Paranoid Schizophrenia on the Basis of the Rorschach - - - 199
Guertin, W. H., and Zilaitis, V. A Transposed Factor Analysis of Paranoid Schizophrenics - - - - - - - - - - - - - - - - - 455

Hammond, K. R. See *Rogers, L. S.*
Hanes, B. Perceptual Learning and Age - - - - - - - - - - - - - - - - - - 222
Heathers, Louise B. See *Merrill, R. M.*
Heineman, C. E. A Forced-Choice Form of the Taylor Anxiety Scale - - - - - - 447
Hess, Judith. See *Cowen, E. L.*
Hiler, E. W. See *Gibby, R. G.*
Holland, G. A. A Comparison of the WISC and Stanford-Binet IQ's of Normal Children - - - - - - - - - - - - - - - - - 147
Holtzman, W. H. See *Calvin, A. D.*
Hovey, H.B. MMPI Profiles and Personality Characteristics - - - - - - - - - - 142
Husni, May. See *Wilmer, H. A.*

Izard, C. E. See *Beier, E. G.*

Jastak, J. Ranking Bellevue Subtest Scores for Diagnostic Purposes - - - - - - - 403
Jenkins, R. L. See *Lorr, M.*
Jourard, S. M. See *Secord, P. F.*

Kantor, R. E., Wallner, J. M., and Winder, C. L. Process and Reactive Schizophrenia - 157
Karlin, L., and Schwartz, M.M. Social and General Intelligence and Performance on the Rosenzweig Picture-Frustration Study - - - - - - - 293
Kaye, D., Kirschner, P., and Mandler, G. The Effect of Test Anxiety on Memory Span in a Group Test Situation - - - - - - - - - - - - - 265
Kenny, D. T., and Bijou, S. W. Ambiguity of Pictures and Extent of Personality Factors in Fantasy Responses - - - - - - - - - - - 283
King, G. F. See *Daston, P. G.*
Kirschner, P. See *Kaye, D.*
Kitzinger, Helen. See *Markwell, E. D., Jr.*
Kobler, F. J., and Stiel, Agnes. The Use of the Rorschach in Involutional Melancholia - - - - - - - - - - - - - - - - 365
Kotkov, B., and Meadow, A. Rorschach Criteria for Predicting Continuation in Individual Psychotherapy - - - - - - - - - - - - - - - 16

Landsman, T. See *Dupont, H. J.*
Lawrence, J. F. See *Williams, H. L.*
Lazarus, R. S. See *Carlson, V. R.*
Lebo, D. The Present Status of Research on Nondirective Play Therapy - - - - - - 177
Levine, J. See *Berkowitz, M.*
Levine, M., and Galanter, E. H. A Note on the "Tree and Trauma" Interpretation in the HTP - - - - - - - - - - - - - 74
Lindgren, H. C., and Robinson, Jacqueline. An Evaluation of Dymond's Test of Insight and Empathy - - - - - - - - - - - - - - - 172
Lonstein, M. See *Rubin, H.*
Lorr, M., and Jenkins, R. L. Three Factors in Parent Behavior - - - - - - - - - 306
Lotsof, E. J. Intelligence, Verbal Fluency, and the Rorschach Test - - - - - - - 21
Lynn, D. B. See *Weisskopf-Joelson, Edith A.*

McConnell, J. See *Calvin, A.*
McNemar, Q. Note on Elwood's Study of IQ Changes - - - - - - - - - - - - 153
Machover, S., and Anderson, Helen J. Validity of a Paper-and-Pencil Form of the MMPI Psychopathic Deviate Scale - - - - - - - - - - - 459
Mainord, W. A. Some Effects of Sodium Amytal on "Deteriorated" Schizophrenics - 54
Mandler, G. See *Kaye, D.*
Maradie, L. J. Productivity on the Rorschach as a Function of Order of Presentation - 32
Markwell, E. D., Jr., Wheeler, W. M., and Kitzinger, Helen. Changes in Wechsler-Bellevue Test Performance following Prefrontal Lobotomy - - - - - 229
Meadow, A. See *Kotkov, B.*
Merrill, R. M., and Heathers, Louise B. A Comparison of the Wechsler-Bellevue and ACE Tests on a University Counseling Center Group - - 63
Miller, D. R. See *Gibby, R. G.*

Ednl. Res. Research
COLLEGE

Mills, E. S. The Madeleine Thomas Completion Stories Test - - - - - - - - - 139
Morgan, H. H. Measuring Achievement Motivation with "Picture Interpretations" - - 289
Mussen, P. H. Differences between the TAT Responses of Negro and White Boys - 373
/ *Muthard, J. E.* The Relative Effectiveness of Larger Units Used in Interview Analysis - 184

Norman, R. D. Sex Differences and Other Aspects of Young Superior Adult Performance on the Wechsler-Bellevue - - - - - - - 411

Pascal, G. R., Swensen, C. H., Feldman, Dorothy A., Cole, Mary E., and Bayard, Jean. Prognostic Criteria in the Case Histories of Hospitalized Mental Patients - - - 163
Peek, R. M. Directionality of Lines in the Bender-Gestalt Test - - - - - - - - 213
Poole, R. C. See *Allen, R. M.*
Pugh, R. W. A Specific Relapse Phenomenon during the Course of Electric Convulsive Therapy - - - - - - - - - - - - - - - - - - 87
Pullen, M. S., and Stagner, R. Rigidity and Shock Therapy of Psychotics: An Experimental Study - - - - - - - - - - - - - - - - 79

Quay, H. See *Sweetland, A.*
Quiroz, F. See *Wattenberg, W. W.*

Rappaport, S. R. Intellectual Deficit in Organics and Schizophrenics - - - - - - - 389
Ray, C. D. See *Allen, R. M.*
Rieman, G. W. The Effectiveness of Rorschach Elements in the Discrimination between Neurotic and Ambulatory Schizophrenic Subjects - - - 25
Robinson, Jacqueline. See *Lindgren, H. C.*
Rogers, L. S., and Hammond, K. R. Prediction of the Results of Therapy by Means of the Rorschach Test - - - - - - - - - - - - - - 8
Rosen, A. Test-Retest Stability of MMPI Scales for a Psychiatric Population - - - 217
Rubin, H., and Lonstein, M. A Cross-Validation of Suggested Rorschach Patterns Associated with Schizophrenia - - - - - - - 371

Schwartz, A. D. See *Feifel, H.*
Schwartz, M. M. See *Karlin, L.*
Secord, P. F., and Jourard, S. M. The Appraisal of Body-Cathexis: Body-Cathexis and the Self - - - - - - - - - - - - - - - 343
Shapiro, Esther. See Graham, E. E.
Smock, C. D. See *Beier, E. G.*
Smith, L. J. See *Granick, S.*
Stagner, R. See *Pullen, M. S.*
Standlee, L. S. The Archimedes Negative Aftereffect as an Indication of Memory Impairment - - - - - - - - - - - - - - - - - - 317
Stanley, J. C. Why Wechsler-Bellevue Full-Scale IQ's Are More Variable than Averages of Verbal and Performance IQ's - - - - - - - - - - 419
Stiel, Agnes. See *Kobler, F. J.*
Stotsky, B. A. See *Gibby, R. G.*
Strother, G. B., and Cook, D. M. Neurocirculatory Reactions and a Group Stress Situation - - - - - - - - - - - - - - - - - - 267
Sweetland, A., and Quay, H. A Note on the *K* Scale of the Minnesota Multiphasic Personality Inventory - - - - - - - - - 314

Swensen, C. H. See *Pascal, G. R.*

Toman, W. Pause Analysis as a Short Interviewing Technique - - - - - - - - 1

Tougas, R. R. See *Beier, E. G.*

Valentine, M. See *Dupont, H. J.*

Walker, E. L. See *Gibby, R. G.*

Wallner, J. M. See *Kantor, R. E.*

Wattenberg, W. W., and Quiroz, F. Follow-up Study of Ten-Year-Old Boys with Police Records - - - - - - - - - - - - - - - 309

Weisskopf-Joelson, Edith A., and Lynn, D. B. The Effect of Variations in Ambiguity on Projection in the Children's Apperception Test - - 67

Wertheimer, M. On the Supposed Behavioral Correlates of an "Eye" Content Response on the Rorschach - - - - - - - - - - - - - 189

Wheeler, W. M. See *Markwell, E. D., Jr.*

Whitmyre, J. W. The Significance of Artistic Excellence in the Judgment of Adjustment Inferred from Human Figure Drawings - - - - - - 421

Wiener, M. See *Cowen, E. L.*

Williams, H. L., and Lawrence, J. F. Further Investigation of Rorschach Determinants Subjected to Factor Analysis - - - - - - - - - - 261

Wilmer, H. A., and Husni, May. The Use of Sounds in a Projective Test - - - - - 377

Winder, C. L. See *Kantor, R. E.*

Zilaitis, V. See *Guertin, W. H.*

New Books and Tests - - - - - - - - - - - - - - - - - 76, 154, 232, 318, 399, 465

Index to Volume 17 - 469

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Pause Analysis as a Short Interviewing Technique

Walter Toman
Harvard University

The basic rule for the patient in psychoanalytic treatment is to say everything that comes to his mind, without selection [1, p. 23; 2, p. 192]. He is to associate freely, and whenever he stops, whenever he makes a pause, the psychoanalyst is, in principle, permitted to prod the patient. Thus the pauses the patient makes become specific points of interaction, and it seems quite possible that some of the peculiarity of psychoanalytic treatment when compared to any other form of psychotherapy springs from this procedure.

In prodding the patient to express what went on in his mind during the pause, or what idea interfered and made him stop, the analyst risks diverting the patient from the ongoing trend of free association, and at times it might be more important, therapeutically, to prod the patient to go on. At times the interruption by the analyst may be quite welcome to the patient, and the welcome topics are not always the ones psychoanalysis is after. In addition we should realize that it is when the patient pauses, that interpretations by the analyst or even questions leading up to interpretations are made. Seldom is such a question raised or an interpretation started by the therapist while the patient is still talking. Thus with a question referring to anything but the pause, we miss the pause, and with any question referring to the pause, we might miss further associations. It takes considerable skill to handle these intricate "molecular" interactions occurring in psychoanalysis, and this skill, in turn, requires training and experience, and experience and training, ad infinitum.

In 1950 H. Rohracher of the University of Vienna suggested an explanation of pauses in speech. While Johanna Karger started working on the quantitative and classificatory analysis of pauses in reading and speaking [3], I began to labor on the dynamic aspects. The reasoning given above was the starting point. Before I describe my proceedings, I would like to give a brief account of Karger's findings.

Twelve subjects had to read a short text aloud. They were sound-recorded and the pauses were explained afterwards by the subjects. Ninety-three per cent of the pauses were connected with punctuation marks, the rest with disturbances of task-awareness, distraction of attention, difficulties in apperception and conflicts between punctuation marks and personal organization of the text. The longest pauses varied between 0.5 and 1 second, and never exceeded 1.8 seconds. In free speech as contrasted with reading (instruction: "Explain why you chose your field of study or your particular profession.") the longest pauses exceeded 1.8 seconds with every subject. It was found that the faster the subject speaks, the greater the relative and the absolute average length of his or her pauses. The ratio of pauses made at all and pauses which the subjects recognized as pauses and explained was 6:1. The estimate of number of pauses made at all was based on the minimal time interval between two spoken words that a keen observer would just begin to consider a pause. This estimate depended on the rigor of the observer, and my ratio of 54.7:17 reported below does not contradict the former ratio. It only indicates a different standard as to what might possibly be considered a pause, and this standard is not crucial for my experiment. The causes of pauses in free speech were: (*a*) disturbances of composition, (*b*) distractions of attention, (*c*) disturbances in production of ideas, (*d*) pauses due to ordering the material, (*e*) time-consuming will processes, (*f*) disturbances in verbal expression (breathing, fatigue, slips of the tongue, etc.). No further specifications were attempted by J. Karger.

The Experiment

My experiment called for a life history of 5 to 10 minutes length, to be spoken by the subject into the microphone of a sound recorder. The subjects were given a casual half-minute of preparation, during which the experimenter got the sound recorder ready, or pretended to do so. Then the subjects spoke their life history into the microphone. They were asked to take about 5 to 10 minutes, and to make it not too factual. The experimenter was present but did not interfere, not even by nodding or by any changes in facial expression.

After the life history was completed, the experimenter and the subject went over the record and analyzed the pauses. Only then did the subjects learn that the pauses were the object of investigation. About three-fourths of the subjects were surprised by the fact that apparently so much had gone on "between the lines" of their performance that otherwise would have escaped their notice.

Results

All of 35 subjects ranging from 18 to 34 years (13 male and 8 female students, 6 male and 3 female adults in nonacademic professions, and 3 male and 2 female adults in academic professions) revealed content material hidden or suppressed in the pauses that probably could not have been brought to light by other interviewing procedures in the given time, i. e., 5 to 10 minutes talking and 10 to 30 minutes of analyzing the pauses of the monologue. Subsequent to the test an interview was conducted with every subject in order to enable the experimenter to further appraise the significance of the content filled into the pauses by the subject while going over the record with the experimenter. A few examples of the performance will be given below.

The subjects produced life histories of an average length of 675 words, with a standard deviation of 230 words. They made an average of 54.4 pauses (with a standard deviation of 26), of which 17 pauses on the average (with a standard deviation of 9.2) could be interpreted by the subjects other than by mere formal difficulties of phrasing or searching for appropriate words.

In order to obtain some sort of an objective measure for the significance of the content material given by the subjects when interpreting their pauses, they were unexpectedly asked to come for another session 1 to 3 months later. Then they were exposed to their records again and asked to interpret the pauses again as far as they could still remember them. On the average 5.3 pauses were recalled (with a standard deviation of 2.7). The content of the other pauses had been forgotten. It turned out that 2.8 of the recalled pauses, i.e., of the recalled ideas or images or thoughts that had originally interfered with the delivery of their life history, were apparently affect-laden early memories, even very early childhood memories, that had oddly popped into the subjects' minds while speaking into the microphone. The rest were recent memories concerning the subjects' present activities and occupations, recent personal relationships and thoughts pertaining to the performance of the test.

These latter memories were presumably not too difficult to recall nor even to reconstruct from the context; but the earlier memories that would appear out of context of the delivered life history to anyone but the subject himself or a sophisticated psychologist would have been impossible to recall unless the same dynamic factor that had originally brought them up had been reactivated. These memories seem to stand for stronger (older and possibly unconscious) dynamic constellations. We may conclude that they were significant, important memories, otherwise the subject would not have recalled them on the right spot in the objectively inconspicuous context of his delivered life history. In that sense this check may be considered an objective measure for the significance of the content material given by the subjects when interpreting their pauses.

Four examples of records interpreted immediately after the test are given below. All further references and comments to the cases have been omitted. They will serve only as a demonstration of the test performance including the interpreted pauses. Thoughts going on during the pauses and reported by the subjects while going over the sound record with the experimenter are underlined.

Examples

Subject A (male, 20 years old)

I remember very little of the first few years. Just that I was born in K. and moved immediately to D. where I lived until I was about 5 years old. I had a very . . . *a little scene in kindergarden. A little girl was crying lustily when having to stay there. I cried too. We made quite a scene* . . . strong, childish attraction for a little blonde girl there, and we have kept writing ever since, and she came to K. last, last month, which was the first date I have had for, during my college experience. This particularly bothers me. Anyway, then we moved to E. I went to the local grammar school, very . . . *out in the country, worst teachers, poor reputation. Father worked to set up schools. Had many enemies. Mother has practically gone blind (over all this). She is very introverted and sensitive* . . . poor grammar school, actually, and there I had three good boy friends, and one pretty blonde girl again with whom I spent a number . . . *her older brother got between us. Had a tent. Wanted his sister and me to undress and play with each other. I never liked the fellow* . . . of the first few years. I can remember that I used to be terrifically sensitive in grammar school, particularly about playing sports. I was very slow on my feet. I could not run in the right way. I used to run on my toes instead of on my feet. So the others would goad me into playing tag with them, practically every day during the first 6 years of grammar school, it seems, anyway. And I would get . . . *I got very hurt and nervous. My best friends were very callous. Were best friends, and stabbed you in the next minute. I was worried about other people's opinion about me.* . . . I tried not to show my anger and, but it did affect me very much. I was very easily offended and hurt. I remember one time that one boy who was actually my best friend made a nasty remark, and I had to stay home that afternoon, because I could not. . . . *I lowered myself by offering him a bicycle ride home. Felt disgusted with myself* . . . face going back to school again. Then going to high school I did very well in school. I was number one student in town graduating to high school and so I had a very easy time. My brother went to high school the year before, and he made such a terrifically good impression that I got by on his own scholastic ability, that is, they gave me all A's without really knowing. . . . *I felt inferior to my brother. But then I became the first in town. My brother had left town* . . . what my own talents were. Although I don't know whether I would have gotten A's or not. Probably yes. But I never had a date with a girl during high school and this particularly bothered me. I never learned to dance, and I felt that was one of the most unfortunate parts of it. . . . *My parents should have encouraged me. They let me down. Brought me up in the wrong way. Later on I talked it over with my father. He said: "We never stopped you." But I had the impression, they did not approve.* . . . My parents during this time never encouraged me or inhibited me from going with girls, but I never had the courage to do it myself, and it has carried over until now, practically. I suppose this is the most interesting part of what I am supposed to say. I don't know. And then I wanted to get on the football team in high school, but never did actually try out for it. Perhaps I could have made it. Then during my junior year of high school my parents decided they wanted me to go to A. for my senior year, and so I had to work very, very hard then in order to make the scholastic standing necessary to get into A. I did get into A. with full scholarship and felt pretty good about my own capabilities then. But when I got there I found that in the first term in A. I got a C minus-and-a-half average, which was very, very poor, particularly for a scholarship boy, so after that I became a recluse and spent about all of my time studying . . . *there were dances in A., Saturday nights. I wanted to go, but did not. Used to sit in my room* . . . so that I had no social life to speak of at A. I never joined a club or anything. I just stayed in my room and ground away. Then I . . . *My brother was in college. I was just going to follow him. I was not very ambitious to go. I just as soon would have worked* . . . did not give much thought to college. I did not really expect I would be able to go to college, because of financial reasons, but I half-heartedly applied, took the college entrance exams, and U. admitted me. It was the only college I had applied to. And . . . *they should have (awarded me a scholarship). I had good records, eventually. A. did not elect me for the cum laude society. They slighted me. They sent me to Washington, free, for a week, though. My brother had been elected.* . . they did not award me a scholarship, though. The only way I came here was that A. managed, A. did give me a prize scholarship, $400, and also the X.Y. Association gave me a scholarship. And also a very old lady who just died gave me a gift of $2500 for my school, my college career, which I could not have done without, and I did finally get to U. . . . *money stretched over 4 years made things easier, but still pretty hard.* . . . And during my freshman year in U. I roomed with a Jewish boy who was about as introverted as I was, and I found myself not particularly liking him, but pitying him in a way. . . . *in French class two girls were sitting next to us, but we never approached them* . . . and we became rather good friends. It was a rather artificial friendship, since I . . . *this amounts to saying, I am superior. His father (who had invented something and discussed advertising it with his son) said: "We put a nice nude on it." The boy had some nasty habits, was dirty, always picking his nose* . . . pitying him, as I did, it was a rather superior feeling I had towards him. Nevertheless we did become very good friends. We, we both wanted to go out with V. girls, but we never managed it. I managed to rationalize myself into thinking that it was because of my low finances, and that's partly true, I suppose . . . *it was in me, of course* . . . but the result was what I figured a rather poor sexual adjustment. And being a senior now this has carried over. And I have had one date in my life time, so to speak, that is a couple during . . . *my and her parents were along. She was older than I was* . . . during the summer times, picnics and that sort of things, but I find it very, very difficult to approach girls in a. . . . *I am getting all worked up. This problem is a contemporary one* . . . in a way that's not self-conscious.

Subject B (female, 19 years old)

I was born in N., and I can't remember anything until I was about 3 years old . . . *memory of the house we used to live in: picture of my uncle standing in the toilet; father said: "He is taking a picture. Come on!"* . . . and I can remember being a happy child. I think I got a lot of attention. . . . *I was over-protected, I guess* . . . from my parents when I was very young. And I had a lot of success in school, and even in kindergarden I remember I was the best reader . . . *from when on was I a lucky child? What facts do I have?* . . . and I could draw and paint . . . *this is not enough. What did I do at home? Can't find anything* . . . and all that sort of thing. And I went to elementary school. In about the third grade, I suppose, I was, oh, 9 years old, I had a little trouble with the children in my class. I was much more withdrawn. I was more imaginative, introverted than they were. I would not really understand the more social gang-like times that they had. . . . *I felt left out. I used to cry. Mother was so helpful. She handled me so good* . . . and I felt left out. I felt very unhappy. I remember they used to make fun of me. And I was always the kind of kid that used to go by herself and, oh . . . *fairies, later on love, also nature, lose myself in those things. Reading and writing poems. Thoughts like: How do you feel the infinite? What happens if you fall forever?* . . . imagine things a lot. Oh, I wrote poetry starting from when I was 6 years old, I think. And I went to . . . *much earlier than high school I became a social being. I left out girls, got along very well with boys, was never in a gang. The girls used lipstick, they were teen-agy quite early* . . . a very, very fine high school in N. I always had a lot of academic success, and in spite of this early withdrawnness of mine, I guess, I was fairly popular, and in high school I majored in art, and I became interested in dancing and music and all these other things, and I had a wonderful time, a lot of social success, academic success. I became good in many, many things, and I really, really enjoyed school aaa . . . *so much to tell. People were so congenial, liberated, very alive and lively atmosphere, good relationships with teachers* . . . a great deal. As far as my home life went at that time, I think, starting around, oh, 15, I began. . . . *I wanted to probe everything. Not deny any experience. Look for the truth.* . . . I began to rebel quite strongly against my family, because I did not think that what they demanded from me was what I wanted to fulfill. And I feel now, that they were very, very . . . *there was a fight. I did a lot of square dancing. Wanted to go with a Negro boy. They forbid it, but I did it. There was hell in the house* . . . oh, they expected a lot in the way of nominal achievement and following what their goals were, I think, much too much. And starting at 16 I began to rebel. I became critical of American culture. I think, yes, and this was also the same time when I thought I was discovering that I was popular and that I had a nice personality, and that the boys liked me, and that sort of thing. And. . . *nowadays I dislike my adolescence. So much dishonesty in that age, effort to make impression, feminine wiles. Disgusting* . . . this went to my head for a little while, and after that I became, as I said, critical and went to the other extreme. I joined a, aaa . . . *how should I characterize it, express the spirit?* . . . youth Zionist group which was very, very, very revolutionary in its mores and, oh, no lipstick and girls treated completely equal with men. . . . *These were the goals, but they did not really achieve them. Girls became masculine. No positive culture of their own* . . . oh, really a little community within itself, and Marxist orientation . . . *they were not even good Marxists* . . . and tried very hard to establish honesty and introspection, and complete naturalness . . . *some of the girls opposed integration, thought Zionism is the solution, quite doctrinaire* . . . I guess, as goals. And oh, when I came to C. as a freshman, I was just, just coming out of this. . . . *Girls in C. less accepting than in N., pressure toward conformity; it was hard to accept them as people. I had a real war for a year.* . . . I was always somewhat critical, and yet I found things there I felt, and still feel, . . . *competition, selfishness, sterility of the whole culture* . . . are lacking in American culture. And I became very interested in Judaism as a way of life, and politically a Zionist, and, well. . . . *Oh, I am up to the present* . . . at this point I no longer am a political Zionist. I am very interested in Judaism. I am interested in modern dancing. I do more of that, practically, than anything else. I have made a conscious effort to narrow down my . . . *talents? That's conceited* . . . fields of expression and chosen dancing as the one I more or less concentrate on. I hope after college. . . *well, I don't really know. It's all vague* . . . somehow, oh yes, I have always been interested in. . . . *I left out my settlement experiences. Do I really have abilities in therapy?* . . . therapy and helping people. I did a lot of work in settlement camps during the summers with a group of kids. I like to teach very much. I enjoy working with children, and I would like to combine therapy with dancing. And possibly go to Israel after college. I guess that's about all.

Subject C (male, 27 years old)

. . . *Shall I tell him my mother was psychotic, or not?* . . . . Well, I think, one of the things that's important in my life history is the. . . . *I am going to tell him that, but I start somewhere else* . . . is the, oh, type of place of my childhood. I grew up in a very small town in the West, in a culture quite different from the one that I am in now. . . . *I felt better in childhood than now. Is that neurotic?* . . . aaa, life there is quite in contrast to the one life one leads around here. It's a rather quiet, free and easy sort of existence, whereas everything here seems to me to be very hostile, full of pressure, aaa. . . . *What shall I say next?* . . . to accomplish things all the time. I find considerable lack of time to sit back and relax and sort of give in. . . . *I better not get emotional* . . . to the enjoyment of . . . *existence? living? that's a poor sentence* . . . my existence. Aaa, another very important thing in my life are my parents, aaa. . . . *Whom shall I take first?* . . . . My father was a very quiet . . . *image of him* . . . retiring sort of person, very, very conscientious, working extremely hard, and not very . . . (*some thought that subject could not recall*) . . . had not very much to say when he was at home. Seems to me, he was a sort of a shadowy background figure in the household rather . . . *but he was warm, even though reticent* . . . than a dominant person. My mother . . . *her father very stern, militaristic, had a lot of control, her affective needs starved* . . . grew up on a farm in the West, had, I think, a rather hard childhood herself that carried over to her later life, and . . . *unfortunately, although I should not say so* . . . to my life. I think, she tried. . . . *I better try to understand* . . . overly hard to compensate for some of the things that she missed in her childhood, and as the result of that she tried too hard to make things easy for myself and for my brother . . . *image of myself as a child of 6 or 8* . . . tried too hard to . . . *control* . . . protect us from the vicissitudes of life, rather had a certain, had a negative effect on my life. I think I lack a lot of . . . *image of a very energetic, goal-directed person* (*male*) . . . drive and autonomy that I might otherwise have had . . . *some sort of therapy might help, I hope* . . . aaa. I think, probably the most significant thing in my life that I can think about, if I were to point to a single incidence, was the psychotic break that my mother had when she was about etc.

Subject D (female, 21 years old)

The first thing I remember is wanting to go to school and starting a little early, and liking it very much . . . *picture of myself, over-confident, boastful* . . . and all during grammar school I was pretty sure of myself and liked it. And when the first . . . *my father did not want me to go to school. My sister went. It was too hard. I was not strong enough, they said* . . . conflict came, when it became time to choose a high school; I chose one that was strictly college preparatory, to which girls from all over R. went, as opposed to a local high school. And as soon as I started that I lost confidence . . . *picture of myself. What would it be to others, when I make new friends?* . . . felt more pressure upon me. I had to work harder, was not sure of myself as far as making friends went, and although I liked school very much, I disliked the teachers, the pressure I felt from them. . . . (*After 10 seconds*) *I hate not to be able to say anything* (*further 20 seconds elapse*). . . . It was an all girls' school which was nice from the point of view of having girl friends, though we were always quite shy about meeting with boys and did not like to go for that reason. And the school put a lot of pressure on it, towards choosing a good college, towards making us want to go, but never gave us a chance to decide whether or not we wanted to go. It just assumed that we would. And upon going to college I have never again felt as much pressure as I did in high school and was glad to go to a college that is at least partially coed. I also felt that the situation was a lot more natural.

Discussion

There seem to be certain advantages to this technique. One of them is the following: The subjects tend to organize their thoughts as well as possible and to give a somewhat coherent story. The fact that their performance is sound-recorded increases their concern with the quality of their performance. On the other hand, they are not given sufficient time to organize their story in advance. Rather they are caught by the sound recorder in the process of organizing, and in this sense the technique may be called a situational test. All criteria developed for appraising verbal creations are eligible and may be applied.

The effort of organizing the story for the sound recorder is presumably greater than the effort of organizing it for a listener in a more familiar situation like interviewing, or counseling, or any conversation. A little speculation may be inserted here, that has been confirmed on other grounds by the author. If we assume that a higher effort put into the performance would consume more "psychic energy," we would expect less energy to be available for any concurrent activities. The main concurrent activity in the delivery of one's life history consists of the defenses against making slips, against revealing too much or uncovering one's emotions more than intended. We would predict that the greater effort may either lower the effectiveness of the defenses against unwanted impulses or, if this is not the case, the defense mechanisms themselves may become more apparent. Both types of effects may give the counselor or therapist cues about the subject that he might not be able to secure otherwise, at least not in the given time or not simultaneously with a life history of the subject. Examples of the first type have been given (subjects A, B, C). Examples of the second type are a few meagre life histories among my records, with long pauses and frequent blanks with respect to the pauses (subject D). These subjects reported that realizing that they made pauses extended one or the other pause to an unusual length, and this in turn blocked their performance, and caused further pauses, until they eventually gave up.

Another advantage becomes clear when we compare this technique with psychoanalytic therapy in particular, but also with any counseling or interviewing technique. As pointed out in the beginning, the pauses are crucial events in psychoanalytic treatment, and however the therapist deals with a single pause, he is interfering. The reported technique yields a record of the uninterrupted cognitive processes of 5 to 10 minutes in length, a sufficient time for a fruitful systematic analysis of the meaning of the pauses.

A third advantage becomes apparent when the record is reviewed. The subjects are surprised by what they said, even surprised by their voice, if they have not heard it recorded before, and this surprise contributes strongly to their willingness to cooperate in the analysis of their pauses. They feel encouraged to take a more objective attitude toward what they said into the microphone than they do when reference is made to what they said in an ordinary interview or conversation with no recording. All subjects except the few who had given very meagre life histories with extremely long pauses were willing to talk very frankly and openly about themselves and their lives, but even the defensive subjects were willing to talk at least about their inhibitions and defenses and found some relief in so doing. Subject and experimenter (client and therapist) are both dealing with the record of the subject's past performance, and this may account for the objective attitude the subject is able to assume quite easily. According to an accepted psychoanalytic metaphor one part of the ego has to cooperate with the analyst in getting a hold on the unconscious parts of the ego (and on the other systems). The pause-analysis technique facilitates this cooperation. The subjects slide into a good contact and rapport with the counselor or therapist, while talking about their performance, their pauses, and even their defenses.

A fourth advantage is the possibility of going over the subject's performance at any later time. What pauses are recalled that have been interpreted (given content and meaning) by the subject in the first session may give additional cues (see "Results" above). Furthermore, the attitude of the subject to his now quite remote performance can be appraised and utilized in several clinical aspects.

Do the subjects manage to interpret all the pauses they make? I would say, they manage to interpret all the pauses that are important as to contents and dynamics. Of 54.4 pauses on the average counted by the experimenter, 17 pauses on the average were commented on and interpreted by the subjects. All the other pauses could safely be considered as pauses due to formal difficulties, difficulties of finding the appropriate words or of finishing a sentence. The subjects were asked to tell what went on during the pauses while hearing their records, and in the beginning the experimenter stopped the record at whatever he considered a pause, but soon the subjects began to cooperate quite efficiently by telling when to stop the record and when not to stop it. Among other reasons, this could be relied on since the subjects had been told that they could refuse to talk about things they did not want to confide. In this case they had only to indicate that there were such things. However, very few subjects resorted to that privilege.

Can we believe what the subject reports? This question may be raised just as well with any psychological approach to personality, and particularly with any form of interviewing. On the grounds mentioned above (good rapport with experimenter, increased objective interest, etc.) we may expect more reliable reports from the subject with the pause-analysis technique than with other short interviewing techniques.

Why apply the analysis of pauses to life histories and not, say, to TAT stories? There is no reason why one should not also apply it successfully to TAT stories. But the life history of a subject is probably the most crucial and important of all the stories he can tell. Any therapeutic effect can, in some sense, be viewed as a recentering, a revision of the life history of the subject, and it is not the events that are revised, but their evaluation. The disappearance of neurotic symptoms is only a secondary effect. Life histories, verbal or written, given by the patients before and after treatment may in general be the most valuable and easy-to-obtain, although certainly imperfect objective checks on these revised evaluations, on a changed self concept, on a liberated and strengthened ego, or on whatever else a therapeutic success may be said to be based. We may also maintain that life histories can be produced repeatedly with interferences by previous performances of no greater intricacy than we would find with other complex clinical tests.

All the shortcomings of this method, such as reliance on what the subject himself tells about the pauses, too many variables, too little quantification, no complete elimination of social interaction and of its interference with the performance to uncontrolled degrees, etc. hold, in one way or another, for any clinical test. In the pause-analysis technique the emphasis is on contents and dynamics and their clinical appraisal. Tendencies pertaining to the person's (subject's, patient's) specific structure and organization of the mind, but also to his performance and to the test situation can be traced via contents of the pauses and its placement in the context of the delivered life history, and any information that could not have been secured by other techniques in the same length of time will be valuable in clinical practice. The technique seems particularly promising in cases on which relatively little time can be spent.

Summary

A clinical technique has been described in which the subject is asked to tell his life history into the microphone for 5 to 10 minutes. After that the subject and the experimenter (counselor, therapist) go over the record and analyze the pauses the subject made. Some quantitative results of 35 subjects have been mentioned and four examples given. Theoretical implications, advantages and disadvantages of this technique have been discussed.

Received June 23, 1952.

References

1. Fenichel. O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.
2. Freud, S. The interpretation of dreams. In *The basic writings of Sigmund Freud.* New York: Modern Library, 1938.
3. Karger, Johanna. Psychologische Ursachen von Sprechpausen. Unpublished doctor's dissertation, Univer. of Vienna, 1952.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Prediction of the Results of Therapy by Means of the Rorschach Test[1]

Lawrence S. Rogers
Veterans Administration Mental Hygiene Unit, Denver, Colorado and University of Colorado

and Kenneth R. Hammond[2]
University of Colorado

One important problem facing mental hygiene clinics is attempting to evaluate before beginning therapy whether or not the patients have the capacity to profit from psychotherapy. Much time and energy ordinarily devoted to the unsuccessful treatment of cases might be more profitably diverted to the better therapy risks should it be possible to make such an evaluation correctly. Our problem, therefore, was to investigate the possibility of using the Rorschach test for prognostic purposes. In short, can the Rorschach test be used to predict which cases will be successful in therapy in a mental hygiene clinic?

Method

The method used involved establishing criterion groups: one group composed of patients who were successful in psychotherapy; the other group composed of patients who were unsuccessful in psychotherapy. The first step was to proceed *deductively* to differentiate between Rorschach protocols of patients in the criterion groups. The next step was to proceed *inductively* to discover whatever differences might exist between the criterion groups and then test these differences on a new sample.

Our deductive procedures included first, a "total impression" judgment. That is, psychologists read the patient's Rorschach protocol and predicted whether the patient was a member of the successful or unsuccessful group. Secondly, "signs" derived by Harris and Christianson [2] were utilized to assign the Rorschach protocol to either criterion group. Thirdly, predictions toward the criterion were made on the basis of whether the *M* present in the protocol was of the flexor or extensor type. Fourthly, two different combinations of Rorschach variables thought likely to discriminate between the criterion groups were tried. The above four procedures, then, were a priori.

Our inductive procedures followed the usual pattern of ascertaining differences between "known groups." First, various Rorschach scores or variables were compared in the two groups. This comparison was made in terms of the percentage of persons in the criterion groups having a given score on a given variable. Secondly, patterns were developed from an analysis of variables which seemed promising as discriminators and these patterns were applied to a new sample. The above two procedures, then, were a posteriori.

Procedures and Results

The procedure began with the establishment of the two criterion groups. One group consisted of 50 protocols of patients whose case file showed that at termination they were thought "unimproved." The other group consisted of 59 protocols of patients terminated "slightly improved" or "improved." The protocols were rescored according to Klopfer's method — using, however, Beck's tables of $F+$ and popular responses. All cases were selected at random from the clinic files, without regard to diagnosis or therapist. All subjects in the criterion groups, as well as in the cross-validation groups, were veterans who had had treatment at a Veterans Administration Mental Hygiene Clinic.

---

[1] Reviewed in the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the authors are the results of their own study and do not necessarily reflect the opinion or policy of the Veterans Administration. This paper was read at the 1951 meeting of the American Psychological Association.

[2] Also contributing to this research were Lewis Bernstein, David Ficks, Lillian Sandberg, and Joanne Knauss.

The judgments concerning the success of the patients in psychotherapy were made by the patients' therapist in consultation with a senior psychiatrist. The senior psychiatrist was one who had been regularly consulted on the case. There were eight social workers, eight psychiatrists, and three psychologists involved.

Deductive Procedures.

The deductive procedures were as follows: *Total impression judgments.* Three psychologists, with Rorschach experience ranging from three to five years, judged each Rorschach protocol for success or nonsuccess under four different conditions — each condition separated by at least a one-month interval.

The procedure under Condition I was to make a *snap* judgment, in not more than two or three minutes, concerning success. The procedure under Condition II was to make a *studied* judgment, with as much time allowed as was felt necessary by the psychologist making the judgment. Condition III was the same as Condition II except that the name of the patient's therapist was given to the judge as additional data. Condition IV provided a check for Condition III in that the therapist's name was given to the judge without the Rorschach protocol. That is, predictions of success were made from the therapist's name alone.

Validity coefficients were calculated for each psychologist under each condition and were expressed in terms of chi square computed for a four-fold table.[3] All results from the total impression judgments were negative. Predictions were no better than chance under all four conditions. Results are presented in Table 1.

Table 1

Validity of Total Impression Judgments

Psychologist		Type of Judgment		
		Snap	Studied	Therapist
A				
	χ^2	.151	.054	.400
	p	$>.60$	$>.80$	$>.50$
B				
	χ^2	1.166	.156	.516
	p	$>.25$	$>.60$	$>.40$
C				
	χ^2	1.37	1.22	.035
	p	$>.20$	$>.20$	$>.80$

Interpsychologist and intrapsychologist agreement under each condition was also calculated and was expressed in terms of tetra-

Table 2

Tetrachoric r's Indicating Inter- and Intrapsychologist Agreement on Total Impression Judgments

(Italicized numbers refer to intrapsychologist agreement)

	Psychologist								
	A			B			C		
	Snap	Studied	Know thpst.	Snap	Studied	Know thpst.	Snap	Studied	Know thpst.
A									
Snap		*.75*	*.74*	.44			.30		
Studied			*.87*		.50			.21	
Know thpst.						.60			.32
B									
Snap					*.65*	*.65*	.51		
Studied						*.53*		.46	
Know thpst.									.27
C									
Snap								*.85*	*.80*
Studied									*.85*
Know thpst.									

---

[3] An objection to the use of chi square might be raised here on grounds that the judgment of one case was not independent of the judgment of another. An attempt to meet this objection was made by not telling the psychologist how many cases were in each criterion group and warning him that he was *not* to assume that the criterion groups were nearly equal in number. Thus, it was hoped that the psychologist judged each new case independently of how he had judged previous ones.

choric *r*. Tetrachoric *r*'s[4] for the judgments between different psychologists ranged from +.21 to +.60 and are presented in Table 2. Correlations above +.40 are significant at the 5 per cent level. Five of the correlations range above +.40 and they are to be considered significant. According to Wilkinson's table [3] the probability of obtaining five correlations significant at the 5 per cent level by chance when nine are calculated is less than .0006.

Tetrachoric *r*'s for the same psychologist under Conditions I — III ranged from +.53 to +.87 and are presented in Table 2. All of these correlations are significant at the 5 per cent level, although how contaminated they are from judgment to judgment is unknown. The judges expressed the not unlikely point of view that contamination from condition to condition was minimal because of the large number of cases (109) and the time interval.

Harris and Christianson's scoring method. Weights for various scores as derived by Harris and Christianson [2] in their study on the use of the Rorschach in the prognosis of therapy were applied to our data. This procedure did not discriminate between the criterion groups ($\chi^2 = .110$, $p = .75$).

M responses. An analysis was made of all the *M* responses with regard to extensor or flexor quality. It should be noted that the examiners did not press the subjects for the direction of movement; any such expression was completely unsolicited. *M*'s appeared in 76 records; in 30 cases there were only flexor *M*'s, in 13 cases only extensor, in 12 cases there were both, and in 21 cases the direction of movement could not be judged. Of the 30 cases in which there were only flexor *M*'s, 12 were improved cases, and 18 unimproved; of the 13 cases in which only extensor *M*'s appeared, 11 were improved and two unimproved. The presence of extensor *M*'s only, therefore, appears to be associated with improvement in therapy. These data seemed promising and are elaborated below.

Response patterns. An attempt to differentiate between the criterion groups on the basis of different patterns of responses was made. Four variables were chosen, not because in combination they represented a particular personality structure, but because of their importance as Rorschach variables. These variables were *R*, *M*, *F*%, and ΣC. A second procedure of this sort was made with five variables, *R*, *M*, *F*%, number of rejections, and presence or absence of additional responses. The second procedure differs from the first in that the last two variables mentioned were included because they were regarded as useful prognostic signs. The hypothesis was that those patients with fewer rejections and a greater number of additionals would be the better therapeutic risks.

A tabulation method was used to analyze the data. Space does not permit complete presentation of this method here. See Hammond [1] for a detailed description of the method.

In the first procedure each variable was broken into categories of High, Medium, and Low.[5] For example, Low $R = 0 - 15$, Medium $R = 16 - 26$, High $R = 26+$. For Low and Medium categories of *R*, *M* was categorized as follows: Low $M = 0$, Medium $M = 1$, High $M = 2+$. For the category of High *R*, Low $M = 2$ or less, Medium $M = 3 - 5$. High $M = 6+$. *F*% was categorized as follows: Low $F\% = 0 - 40\%$, Medium *F*% as 41 - 60%, and High *F*% as 61%+. ΣC was categorized in this fashion: For Low ranges of *R*, Low $\Sigma C = 0$, Medium $\Sigma C = .5 - 1.4$, High $\Sigma C = 1.5+$. For Medium ranges of *R*, Low $\Sigma C = 0 - 1.0$, Medium $\Sigma C = 1.5 - 2.4$, High $\Sigma C = 2.5+$. For High ranges of *R*, Low $\Sigma C = 0 - 1.9$, Medium $\Sigma C = 2.0 - 3.4$, High $\Sigma C = 3.5+$.

For the second procedure, *R*, *M*, and *F*% remained as described above. Rejections were categorized in the same way for all values of *R*, Low rejections = 0, Medium rejections = 1, High rejections = 2+. "Additionals" were categorized merely in terms of presence or absence. Results from the pattern analysis procedure were negative. By inspection it was

---

[4] Splits closely approximate .50 — .50. Condition IV is omitted.

[5] Decisions about the range in categories were made from an examination of the data obtained. Although recognized as a risky procedure, it appeared to be the most promising one and risks would be forestalled by use of a cross-validation group, if positive results were obtained.

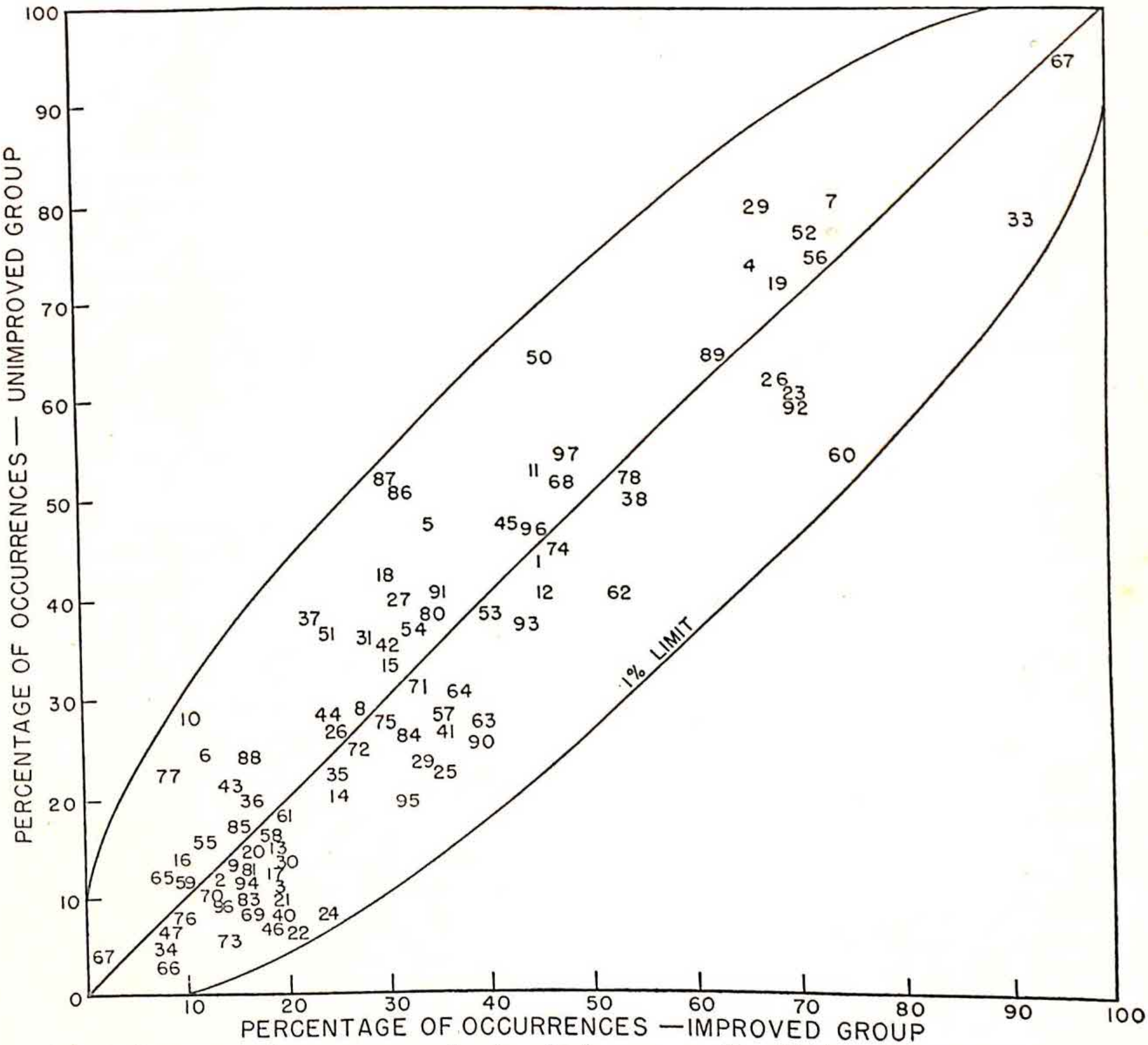


Fig. 1. Percentage of responses occurring in criterion groups. (See Table 5 for responses indicated by numbers in this figure, e.g., $1 = R < 20$.)

obvious that the patterns used did not discriminate between the criterion groups.

In summary, then, only one a priori method of prediction, type of *M*, seemed even promising; the other methods offered no evidence of better-than-chance success.

Inductive Procedures

The inductive procedures were as follows: *Single variables:* The first step was the analysis of single variables, or scores, such as *M*, *ΣC*, *F%*, etc. Each such variable was examined to discover whether the percentage of persons achieving a certain score differed in the criterion groups. For example, the percentage of persons in the criterion groups who achieved three or more *M*'s was compared. The procedure for choosing cutting points was deliberately that of choosing those cutting points which seemed empirically most promising. There were two reasons for this procedure: (*a*) The study had now become a frankly empirical one and cross-validation was planned, and (*b*) it very soon became apparent that few large differences were going to be found.

Ninety-nine percentages were compared and are presented pictorially in Fig. 1 by means of a method developed by Zubin [4]. It will be noted from the figure that the results are negative. None of the variables produced significant differences between the criterion groups.[6] Table 3 gives a list of the signs tested, corresponding to the numbers in Fig. 1.

Table 3

Rorschach "Signs" Corresponding to Numbers In Figure 1

1. *R* less than 20
2. *R* less than 10
3. *R* 40 or more
4. *M* presence
5. *M* more than 1
6. *M* more than 3
7. *FM* presence
8. *FM* 4 or more
9. *FM* more than 4
10. *m* presence
11. Add. *m* presence
12. *k* presence
13. *k* more than 1
14. *K* presence
15. *FK* presence
16. *FK* 3 or more
17. *F* 4 or more
18. *F* 3 or more
19. *Fc* presence
20. *Fc* 3 or more
21. *c* presence
22. *C′* presence
23. *FC* presence
24. *FC* more than 2
25. *FC* more than 1
26. *CF* presence
27. *CF* more than 1
28. *CF* more than 2
29. $FC > CF$
30. *FC* to 0 *CF*
31. $CF > FC$
32. *CF* to 0 *FC*
33. Presence of *CF* or *FC*
34. Presence of *C*
35. Human absence
36. Human more than 3
37. Human more than 2
38. *Hd* presence
39. *Hd* more than 1
40. *Hd* more than 3
41. $Hd > H$
42. *A%* 50 or more
43. *A%* 60 or more
44. *A%* 30 or less
45. Populars 5 or more
46. Populars 2 or less
47. Populars 8 or more
48. $A > Ad$
49. *Aobj* presence
50. *Obj* presence
51. Others presence
52. Categories more than 6
53. Categories more than 8
54. Rejection presence
55. Cut-off 3 or more
56. *At* presence
57. *At* more than 2
58. *At* more than 3
59. *Sex* presence
60. *N* or *Geog* presence
61. *Art/Des* presence
62. Nature presence
63. *Geog* presence
64. Clouds presence
65. Blood presence
66. Fire presence
67. $AD > A$
68. *F%* more than 70
69. *F%* more than 50
70. *F%* less than 30
71. *F%* 60 or more
72. *F%* less than 40
73. Shading absence
74. Shading 15% or more
75. Shading 20% or more
76. Shading 30% or more
77. $Sum\ C = 0$
78. *Sum C* less than 2
79. *Sum C* more than 4
80. *Sum C* less than 1.5
81. $CF > FC$ & $Sum\ C > 3$
82. *M:Sum C*. Same as *OM/OC*
83. % last 3:24 or less
84. % last 3:30 or less
85. % last 3:45 or more
86. % last 3:40 or more
87. % *C* less than 10
88. % *C* 20 or more
89. % *C* 13 or less
90. $W\% > 50$
91. *W%* 30 or less
92. $W\% < 20$
93. *W%* 45 or more
94. *D%* 30 or less
95. *D%* 40 or less
96. *D%* 65 or more
97. *D%* 41 – 60
98. *S* presence
99. *S* more than 1

Rules for prediction. After examination of Fig. 1, and using the analysis of the *M*'s, rules were set up which seemed most likely to discriminate between the criterion groups. These rules are listed in Table 4. The rules were then applied to a new sample of 104 cases, of whom 70 were improved, and 34 were not. Upon application, it was noted that the rules were not mutually exclusive. It was, therefore, decided that each case would be judged only once. Since it appeared from our original sample that we could predict success more accurately than failure, if a case fell into a category predicting success, that category was the only one used. If a case fell into more than one category, it was put in the one with

Table 4

Prediction Rules

I. Considering type of *M*
  Improvement:
  1. Presence of extensor *M* and $\% C > 10$
  2. Presence of flexor *M*, absence of small *m*, $FC > 2$
  3. Presence of flexor *M*, presence of *CF* or *FC*, $\% C > 10$, absence of object, presence of *N* or *G*
  4. Presence of *M* that is neither *E* or *F*, absence of small *m*, $FC > 2$, $\%$ last $3 < 40$
  5. Presence of *M* that is neither *E* or *F*, and absence of object
  Unimprovement:
  6. Presence of extensor *M*, presence of *CF* or *FC*, $\% C < 10$, presence of object
  7. Presence of flexor *M*, $\%$ last $3 > 40$, except when $FC > 2$
  8. Presence of flexor *M*, $\% C < 10$, *CF* or *FC* present, presence of object, and *N* or *G*
  9. Presence of *M* that is neither *E* or *F*, presence of *CF* or *FC*, $\% C > 10$, presence of object, *N* and *G* absent

II. Without regard to *M*
  Improvement:
  10. No small *m*, $\Sigma C > 0$, $\%$ last $3 < 40$
  11. Presence of *CF* or *FC*, $\% C > 10$
  Unimprovement:
  12. No small *m*, $\Sigma C = 0$
  13. $\% C < 10$, presence of object, and *N* or *G*

---

[6] Only those patients who had at least 5 interviews are included here. Identical procedures were used with the total group with negative results.

which the most favorable results had previously been obtained in the original sample. It should also be noted that many cases do not fall under any of these rules and therefore are not judged. Results for each rule are presented in Table 5.

Table 5
Results of Empirical Pattern Analysis
(First sample $N = 50$ Improved, 59 Unimproved; Second sample $N = 70$ Improved, 34 Unimproved)

Rules*	Times cor.	Times incor.	Not judged	Times cor.	Times incor.	Not judged
Considering type of *M*						
Extensor *M*						
#1	15	3		12	2	
(#6)†	4	1		0	1	
			2			6
Flexor *M*						
#2	2	0		5	0	
#3	3	0		2	0	
(#7)†	4	1		4	2	
(#8)	5	2		1	1	
			11			11
M neither *E* nor *F*						
#4	7	2		8	1	
(#5)†	2	0		3	1	
(#9)	5	1		1	1	
			4			10
Without regard to *M*						
#10	4	3		16	7	
#11	42	23		28	9	
(#12)†	11	5		7	10	
(#13)	11	4		9	7	
			6			11

*See Table 4 for rules. Rules are numbered in accordance with their number in Table 4.

†Numbers in parentheses refer to rules which predict unimprovement and vice versa.

The results of the cross-validation study are described in detail below and are presented in tabular form in Table 6.

Predicting Improvement: (*a*) Empirically derived rules for predicting improvement which include consideration of *M* do appear to predict improvement better than chance (see A, Table 6). (*b*) Empirically derived rules for predicting improvement which do *not* include *M* do *not* quite reach the usual criterion for better-than-chance prediction (see B, Table 6).

Table 6
Validity of Empirical Rules

A.
Pred. Imp. with *M*

	Imp.	Ui.	
With rule	30	4	34
W.O. rule	40	30	70
	70	34	104

$\chi^2 = 10.05$
$p = < .01$

B.
Pred. Imp. W.O. *M*

	Imp.	Ui.	
With rule	44	16	60
W.O. rule	26	18	44
	70	34	104

$\chi^2 = 2.34$
$p = .14$

C.
Pred. Unimp. with *M*

	Imp.	Ui.	
With rule	5	6	11
W.O. rule	65	28	93
	70	34	104

$\chi^2 = 2.66$
$p = .10$

D.
Pred. Unimp. W.O. *M*

	Imp.	Ui.	
With rule	17	16	33
W.O. rule	53	18	71
	70	34	104

$\chi^2 = 5.47$
$p = < .02$

Predicting Unimprovement: (*a*) Empirically derived rules which include the consideration of *M* do *not* quite reach the usual criterion for better-than-chance prediction (see C, Table 6). (*b*) Empirically derived rules which do *not* include consideration of *M do* predict unimprovement better than chance (see D, Table 6). The above statistical presentation is modified below so that the results are made more clinically meaningful. For example, one might ask whether these data indicate when the use of *M* in clinical prediction is warranted.

Rules including M: When attempting to predict improvement using rules which include *M*, predictions could be made in 34 cases (32% of the total); these predictions proved to be correct in 30 instances (see A, Table 6). Apparently, when *M* is present this is a useful procedure. However, when attempting to predict *unimprovement* (using *M*) predictions could be made in only 11 cases (10% of total), 6 of which were correctly predicted (see D, Table 6). This does not appear to be a useful procedure. Thus, the clinical use of *M* seems indicated when predicting *im-*

provement, but not when predicting *unimprovement.*

Rules which do not include M: When attempting to predict improvement using rules which do not include *M,* predictions could be made in 60 cases (58% of total); these predictions proved to be correct in 44 instances (see B, Table 6). This seems to be a clinically useful procedure, for the psychologist was correct about three-fourths of the time. Note, however, that the odds are biased in the psychologists' favor when predicting improvement, for the odds are about 2 to 1 that any given case in the population will be improved. It is for this reason that this rule does not quite reach the usual 5% criterion.

When attempting to predict unimprovement using rules which do not include *M,* the psychologist could predict for 33 cases (31% of the total). The psychologist was correct about 50% of the time (see C, Table 6). This hardly seems to be a useful procedure, even though better-than-chance prediction is achieved. Thus, procedures which exclude *M* from consideration seem useful clinically when predicting improvement only.

In brief, then, these data indicate that the rules for predicting improvement would be *clinically* useful, with or without the use of *M,* and that the rules for predicting unimprovement would not be clinically useful, with or without the use of *M.*[7] The reader should note again that in this population prediction of improvement is much easier than prediction of unimprovement because of the distribution of cases, i.e., there are far more improved cases—thus any prediction of improvement is more likely to be right than wrong.

Discussion

These results point to the fact that single factors on the Rorschach are not sufficient for predictive purposes. However, combination of factors, i.e., empirically determined rules for prediction, may be used with some measure of success for certain cases. The rules tested in this study were the first ones selected, and it is quite probable that they are not the best, nor the only rules which might be derived from the data. For example, shading responses were not considered nor were location responses. Other factors may well be considered. For example, we used a representative clinic sample, without regard to diagnosis or therapist. It is quite possible that if the patient population were stratified with regard to diagnosis and therapist, that greater precision might be obtained. However, the results obtained appear promising and further investigation seems appropriate.

On the other hand, the negative character of the results of our a priori procedures should not be overlooked. In this study, four procedures which might reasonably have been expected to have predictive value, did not approach success. Consider especially the "total impression" results. Here intrapsychologist agreement was higher than interpsychologist agreement. Thus, one might speculate that three rather different methods of "total impression" judgment were used rather consistently without success.

Our conclusion is that while the Rorschach test *may* be used for prognostic purposes successfully by some psychologists, there is emphatic need for further empirical check; indeed, empirical work may make prognosis from the Rorschach explicit and practical.[8]

Summary and Conclusions

1. Rorschach records of 109 unselected veterans who had been in treatment were used to predict outcome of treatment.

2. Psychologists experienced in the use of the Rorschach did not make better-than-chance predicitions of outcome of treatment under three separate methods of studying the Rorschach protocols.

3. Combinations of four variables in one instance, and five variables in another, also

---

[7] While differences in efficiency could be evaluated statistically, the prediction rules hardly seem well enough grounded to warrant such refinement.

[8] After this article was submitted for publication, an article by G. Filmer-Bennett entitled "Prognostic indices in the Rorschach records of hospitalized patients" appeared in *J. abnorm. soc. Psychol.,* 1952, **47,** 502-506, in which the author noted that "The foregoing results failed to reveal any sharp trends in Rorschach performance which were significantly related to outcome of treatment or which compared with those reported in earlier studies."

failed to discriminate between the criterion groups.

4. The results obtained by Harris and Christianson for their group did not predict successfully for this group.

5. There was some evidence that extensor *M*'s, when they appear alone in a record, are associated with improvement in therapy.

6. Ninety-nine Rorschach "scores" were examined, and did not differentiate between the criterion groups.

7. Rules for prediction obtained on an empirical basis, and cross validated, showed definite promise of utility. The rule most useful permitted predictions in about one-third of the cases examined, and showed correct predictions at the 1 per cent level of confidence.

Received July 14, 1952.

References

1. Hammond, K. R. A tabulation method for analyzing combinations of Rorschach scores. *J. clin. Psychol.*, 1951, **7**, 276-279.
2. Harris, R. E., & Christianson, C. Predictions of response to brief psychotherapy. *J. Psychol.*, 1943, **15**, 49-74.
3. Wilkinson, B. A statistical consideration in psychological research. *Psychol. Bull.*, 1951, **48**, 156-158.
4. Zubin, J. Note on a graphic method for determining the significant difference between group frequency. *J. educ. Psychol.*, 1936, **19**, 431-444.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Rorschach Criteria for Predicting Continuation in Individual Psychotherapy[1]

Benjamin Kotkov
Boston Mental Hygiene Unit, Veterans Administration
and Arnold Meadow
University of Buffalo

This study is the second in a series of investigations designed to determine Rorschach criteria for predicting continuation in psychotherapy. The hypothesis for the study was suggested by the results of an initial study of group psychotherapy patients [3]. These results had indicated that a group of neurotic patients who continued in group psychotherapy had, on the average, higher *FC* minus *CF, R,* and *D%* Rorschach scores than a matched group of patients who discontinued group psychotherapy without the consent of the psychotherapist. A weighted combination of the three Rorschach scores derived by a discriminant function analysis permitted an accurate allocation of each patient to a "continuing" and "noncontinuing" group in 69 per cent of the cases. A chi-square test indicated that this percentage of successful allocation was significant beyond the 1 per cent level.

The aim of the present study was to determine at an empirical level whether the three Rorschach variables which were useful in determining continuation in group psychotherapy could predict the same behavior in individual psychotherapy.

Subjects and Procedure

The Rorschach records of twenty-six "continuing" and twenty-six "noncontinuing" patients were selected at random from the test files of the Boston Mental Hygiene Unit, Veterans Administration. A patient was classified as "continuing" if he attended the first nine psychotherapy sessions without severing his relation to the clinic. A patient was classified as "noncontinuing" if he had accepted individual psychotherapy, but ceased keeping appointments without the psychotherapist's consent before the ninth psychotherapy session.[2]

The fifty-two Rorschach records were all those of white, male veteran patients referred for individual psychotherapy. The patient group was a predominantly neurotic population with the exception of a few latent schizophrenics and behavior disorders. The age range was from nineteen to forty years with a mean age of twenty-eight years. The educational level of the group ranged from eight to fifteen years, with a mean of twelve years. There were as many married as single men. The majority were of Catholic faith. The occupations varied, with a fairly even distribution of students, clerks, skilled, semiskilled, and unskilled workmen.

The Rorshach tests had been administered by members of the Psychological staff of the Boston Mental Hygiene Unit according to the standard procedure described by Klopfer [2]. The scoring of each protocol was checked by the senior author to insure uniformity and accuracy.

The design of the experiment provides for two statistical comparisons: (*a*) between the means of the continuing and noncontinuing groups on each of the three Rorschach variables, *FC* minus *CF*, *D%*, and *R*.; and (*b*) between the continuing and noncontinuing groups on a weighted Rorschach score obtained by combining the three variables according to the discriminant function equation derived from the previous group psychotherapy study.

---

[1] Reviewed in the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the authors are the results of their own study and do not necessarily reflect the opinion or policy of the Veterans Administration. The authors wish to thank Dr. Donald H. Bullock, University of Buffalo, for his help in preparing the manuscript for publication.

[2] The same number of sessions attended was used in this study as in the previous group psychotherapy study [3].

Table 1

Means of Continuing and Noncontinuing Psychotherapy Groups for Three Rorschach Variables

Variable	Continuing group	Noncontinuing group
$FC-CF$	+.19	−1.08
$D\%$	53.88%	45.46%
R	28.8	16.46

Results

The means of the three Rorschach variables, $FC-CF$, R, and $D\%$ for the continuing and noncontinuing groups of patients are presented in Table 1. The data indicate that the mean *FC* minus *CF* and *R* scores for the continuing group is higher than the corresponding means for the noncontinuing group. The tests of significance between the means yield a probability level of .01 for the *FC* minus *CF* mean difference and .02 for the *R* mean difference. The difference between the mean *D%*'s of the continuing and noncontinuing groups is not significant.

The results of the present experiment which used individual psychotherapy patients as subjects in general parallel the results obtained in the previous group psychotherapy experiment. The mean *FC* minus *CF* and *R* scores of the continuing groups of patients are significantly higher than the corresponding means for the noncontinuing groups in both experiments. In both experiments the mean *R* scores for the continuing groups are higher than the mean *R* scores for the noncontinuing groups. The test of significance for the individual psychotherapy experiment yielded a much lower probability level. The *p* for the individual psychotherapy experiment was .02; for the group psychotherapy experiment it was .01. The difference between the mean *D%* for the continuing and noncontinuing individual psychotherapy groups was in the opposite direction to that found in the group psychotherapy experiments. In the earlier experiment, however, the difference had only been significant at the .2 level.

The results of the discriminant function analysis performed in the first experiment had yielded the equation

$$z = 00241397x_1 + .00006969x_2 + .00038326x_3$$

in which x_1 is the *FC* minus *CF* score, x_2 is the *D%* score and x_3 the *R* score.[3] In the present experiment a *z* score was accordingly obtained for each patient by substituting his empirical *FC* minus *CF*, *D%* and *R* scores in this equation. The distributions of *z* scores for the continuing and noncontinuing individual psychotherapy groups are presented in Table 2. To facilitate comparison with the results of the previous experiment, the comparable distributions of the continuing and noncontinuing group psychotherapy patients are presented in the same table.

Table 2

Frequency Distribution of *z* Scores for Continuing and Noncontinuing Psychotherapy Groups

z score class intervals	Continuing individual patients	Noncontinuing individual patients	Continuing group patients	Noncontinuing group patients
1. −.00283112 − .00000000	0	3	0	0
2. .00000000 − .00283112	1	3	0	0
3. .00283112 − .00566224	3	4	7	13
4. .00566224 − .00849336	1	3	5	19
5. .00849336 − .01132448	6	8	11	17
6. .01132448 − .01415560	3	4	4	0
7. .01415560 − .01698672	3	0	9	4
8. .01698672 − .01981784	2	1	1	2
9. .01981784 − .02264896	2	0	2	1
10. .02264896 − .02548008	2	0	2	0
11. .02548008 − .02831120	0	0	0	0
12. .02831120 − .03114232	3	0	1	0
Total *N*	26	26	42	56

Table 2 indicates considerable overlap between the continuing and noncontinuing group's distribution of scores. All distributions are skewed towards the lower end of the *z* score range. There is, however, a greater proportion of patients with lower z scores in the noncontinuing groups than in the continuing groups in both the group and individual psychotherapy distributions.

---

[3] The *D%* score was multiplied by the constant 100 to eliminate the decimal point and simplify the calculations.

Table 3

Test of Significance between Mean z scores of Continuing and Noncontinuing Groups for Individual Psychotherapy

	N	Mean z scores	Standard deviations	Standard errors	Critical ratio	p
Noncontinuing group	26	.006878	.00463	.000926	4.02	<.01
Continuing group	26	.014478	.00822	.001645		

The results of the test for the significance between the means of individual psychotherapy continuing and noncontinuing groups are presented in Table 3. Table 3 indicates that the continuing group has a higher mean z score than the noncontinuing group. Since the direction of this difference was postulated by hypothesis, a one-sided significance test is justified and the critical ratio is accordingly significant at the 1 per cent level.

Empirical inspection of the z-score frequency distributions of the group psychotherapy patients indicated that the value $z = .01132448$ represented the cutting point which produced the minimum overlap between the continuing and noncontinuing groups. In the present experiment, if we call all those cases above this point continuing and all those cases below it as noncontinuing, we find that we have predicted 36 correct placements and 16 incorrect ones. To determine whether this result could have been obtained by chance, a chi square was computed. The chi square was 7.69, a value which is significant beyond the 1 per cent level.

Discussion

An evaluation of the results of both experiments suggests the conclusion that the *FC* minus *CF* and *R* variables clearly differentiate continuing and noncontinuing patients whether in group or individual psychotherapy. The second conclusion which emerges from the data is that the *D%* is probably not a useful differentiating index in either type of psychotherapy.

When the three Rorschach variables are combined into one predictive index by a discriminant function analysis, the results for the individual psychotherapy experiment closely parallel those obtained in the group psychotherapy experiments. In both cases, the analysis yielded correct predictions in precisely 69 per cent of the cases.

In the group psychotherapy experiment, the discriminant function equation and cutting point z score were empirically derived. Accordingly, it was expected that the experimental error would reduce the accuracy of prediction in the individual psychotherapy experiment. The striking similarity of the results of the second experiment to those of the first greatly increases the degree of confidence with which the relationship, mathematically expressed by the discriminant function equation, can be accepted. However, future predictions would probably be more accurate if only the *FC-CF* and *R* variables were utilized in the equation.

The experimental results are even more impressive when certain differences between the designs of the two experiments are more closely examined. The type of subjects used in the experiments differed in many respects. In the first experiment 71 per cent of the subjects were obese nonveteran patients, and 29 per cent were neurotic veteran patients. All of the patients of the second experiment were veterans, most of whom were neurotic. Thirty-three per cent of the subjects of the first experiment were over 40 years of age. All of the patients of the second experiment were under 40 years of age. Over 50 per cent of the patients in the first experiment were women. The patients in the second experiment were all men.

Despite the differences in the types of subjects studied, the similarity of the results suggests that at least some of the personality factors which determine continuation in group psychotherapy determine continuation in individual psychotherapy.

Other personality factors not measured by the methods employed in these experiments might indicate some differential reaction to the two types of psychotherapy. Similarly, factors other than those related to personality, such as method of psychotherapy, personality of psychotherapist, group composition, etc., may also be important determinants of whether a patient continues or discontinues psychotherapy.

The problem arises as to how to account for the results of the two experiments on a theoretical level.

It is assumed that the motivation of patients who undertake psychotherapy is ambivalent. On the one hand, they wish to continue psychotherapy in order to relieve their symptoms. On the other hand, they wish to discontinue psychotherapy because of the anxiety aroused by the psychotherapeutic process.

It is further assumed that patients with greater *FC* minus *CF* and *R* scores have a greater capacity to tolerate the anxiety aroused by the psychotherapeutic process and accordingly remain in psychotherapy. Conversely, patients with relatively low *FC* minus *CF* and *R* scores have a lesser capacity to tolerate the anxiety aroused by the psychotherapeutic process and accordingly, discontinue psychotherapy.

There are several facts which support this second assumption. Williams [5] reports a positive correlation between the number of *FC* responses and a subject's capacity to withstand experimentally induced stress. A study by Eichler [1] indicates that experimental stress produces a significant decrement in the *R* variable. It is possible that the patients with low *FC* minus *CF* and *R* Rorschach scores at the beginning of psychotherapy have initially a greater amount of anxiety than the patients with higher *FC* minus *CF* and *R* scores. They are accordingly relatively unable to tolerate the additional amount of anxiety induced by the psychotherapeutic process.

Rogers, Knauss, and Hammond [4] have recently published the results of a study similar to the one reported here. These authors attempted to determine whether the Rorschach test would aid in predicting whether an outpatient would return to the clinic for five or more interviews. On the basis of their results, the authors conclude that this behavior cannot be predicted by the Rorschach test, although they qualify their conclusion with the proviso that future, more extensively controlled studies might yield positive results. The contrasting results of the study of Rogers, *et. al.*, and those of the present study, require some theoretical interpretation.

A possible explanation can be formulated in terms of the differences in the methodology employed in the two studies. Rogers, *et. al.* used two methods for analyzing their Rorschach data: (*a*) clinical impression judgments of the entire Rorschach protocol, and (*b*) statistical comparisons of ninety-nine independent scores or signs. An objection to the use of the first method is that there was insufficient reliable information about the relationship of the independent variables (the Rorschach factors) to the dependent variable (continuation in therapy) to permit the formulation of a fruitful hypothesis. This judgment is admittedly a "Monday morning quarterback" type of criticism, but it is at least supported by the results of the Rogers study.

The sign method involves similar difficulties. The cutting points for each of the "signs" are derived on an a priori basis. The previously suggested point, that the present area of investigation is a relatively unexplored one, may also be considered as an objection to the use of a priori cutting points. In addition, the sign method is subject to the criticism that it violates the Gestalt quality of the classical Rorschach interpretive procedure.

The discriminant-function method employed in the present series of studies preserves the advantage of a method which provides for the simultaneous consideration of several Rorschach factors without sacrificing the objectivity afforded by precise quantification.

Summary

Results of a previous study of group psychotherapy patients suggested the hypothesis that continuation or noncontinuation in psychotherapy was a function of three Rorschach factors, *FC* minus *CF*, *D%*, and *R*. The aim of the present study was to test this hypothesis on individual psychotherapy patients.

Two methods were employed. The first method consisted of comparing the means on each of the three Rorschach variables of a

group of patients who discontinued individual psychotherapy with the means on the same variables of a group of patients who continued individual psychotherapy. The second method consisted of comparing the same continuing and noncontinuing groups of patients on a weighted Rorschach index. This index was obtained by the solution of a discriminant function equation empirically derived from the previous investigation of group psychotherapy patients. Twenty-six continuing and 26 noncontinuing clinic outpatients were used as subjects. The following results were obtained:

1. The average *FC* minus *CF* and *R* scores of the continuing patients were significantly higher than those of the noncontinuing patients.
2. The difference between the two groups of patients on the *D%* variable was not significant.
3. The discriminant-function-weighted Rorschach index yielded a correct prediction in 36 out of 52 cases. A chi-square test of the statistical significance of this prediction resulted in a probability value of less than 1 per cent.

The results are interpreted in terms of an hypothesis which postulates that patients with relatively high *FC* minus *CF* and *R* Rorschach scores have a greater capacity to tolerate the anxiety aroused by the psychotherapeutic process than patients with relatively low *FC* minus *CF* and *R* Rorschach scores.

Received July 22, 1952.

References

1. Eichler, R. M. Experimental stress and alleged Rorschach indices of anxiety. *J. abnorm. soc. Psychol.*, 1951, **46**, 344-355.
2. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1942.
3. Kotkov, B., & Meadow, A. Rorschach criteria for continuing group psychotherapy. *Int. J. Group Psychother.*, in press.
4. Rogers, L. S., Knauss, Joan, & Hammond, K. R. Predicting continuation in therapy by means of the Rorschach test. *J. consult. Psychol.*, 1951, **15**, 368-371.
5. Williams, M. An experimental study of intellectual control under stress, and associated Rorschach factors. *J. consult. Psychol.*, 1947, **11**, 21-29.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

11·8.70
J 91

Intelligence, Verbal Fluency, and the Rorschach Test[1]

Erwin J. Lotsof
Ohio State University

It is implied in the literature that some Rorschach scoring categories have relatively distinct meanings: that is, they are related to independent personality variables. However, there is relatively little experimental work on the demonstration of this univocality. Hypotheses often engendered from a Rorschach protocol seldom can be tested or perhaps more important, they do not conform to the formal conditions of testing hypotheses. The terms in the Rorschach are rather vague, ambiguous, and almost defy being operationally defined; one need only examine words like "libidinal urges," "central anxiety," etc. "The problem facing the scientific psychologist is not, is the Rorschach valid, but, what aspects of the Rorschach are valid, and what are its limitations and uses?" [5, p. 306]. Wittenborn [9] using a factorial technique has shown that there is in fact much overlap in the scoring categories of the Rorschach using a college population. In an unpublished paper, Schroder [8] also has found much overlap in the Rorschach categories in a sample composed of individuals having Binet IQ's of 60-80 and CA's of 11.5 to 13.0.

It would appear on the basis of traditional Rorschach interpretations that scores are unrelated to the verbal fluency of the subject. From the point of view taken in this paper and implied by others [3] such a positive relation is expected to hold. In the clinical situation, one often sees the "verbal" individual produce large numbers of responses and with it a rather large number of determinants. The question then arises do these determinants actually give us the "core" of the personality or are they reflections of the patient's ability to verbalize. It was the hypothesis of this study that an explanation of the obtained overlap might be found in the possible relationship of the scoring categories to general verbal fluency. The Rorschach protocol is after all a sample of the subject's verbal repertoire. For example, it is virtually impossible to score for movement if the subject does not employ an action term in his response; likewise the scoring of shading implies the presence of certain descriptive terms in the protocol. If it were demonstrated that various Rorschach categories are not independent of each other, and if it were demonstrated at the same time that they are related to external criteria of verbal facility, then one would be tempted to question the validity of certain Rorschach interpretations. Factor analysis appeared to be the most useful technique for approaching this problem. If, as some Rorschach workers claim, the several determinants are independent, they should manifest themselves as isolated factors in a factor analysis.

Subjects and Procedure

A group of thirty college students, male and female, stratified on intelligence as measured by the Ohio State Psychological Examination, participated in this investigation. Three students were selected randomly from each decile of the OSPE. None of the subjects had taken the Rorschach prior to the present situation. The experimental situation consisted of two parts, both of which were recorded by the use of a concealed microphone. In Part 1, the subjects gave verbal descriptions of five hypothetical situations. In Part 2, each subject was administered the Rorschach test.

---

[1] Read in part at the Midwestern Psychological Association meetings 1951. The writer wishes to thank Dr. J. B. Rotter and Dr. R. Wherry for their aid in this investigation.

The hypothetical situations used are as follows:

1. What are the two main streets which intersect in your town? Just imagine that you are back there now on the corner. It is 12 o'clock noon on Saturday. Tell me at length everything you see and what is happening.

2. I would like you to tell me as much as you can as to what you do for recreation. Tell me all about it.

3. Suppose that you have not had a vacation for a long period of time and one day someone gives you all the money that you need to go on a trip. Tell me at length all about the trip that you would take.

4. Do you know where the oval is? Well, just imagine that you are in the center of the oval and you are telling some friend from home about *everything* you see. What would you tell them.

5. What is the easiest thing for you to talk about? Tell me as fully as possible all about it or a modification of this to fit the subject's topic.

At the termination of Part 1, the subject was told that the experimenter wanted him to do something different. At this time the Rorschach was administered employing the instructions advocated by Beck. In order to minimize the possibility of experimenter bias influencing the number of determinants, a rigid inquiry technique was employed. After the free association period, the examiner asked the following questions for each response produced:

1. Show me the - - - - - - ?
2. What gave you the idea that it was a - - - -?
3. Can you tell me anything else that suggested that it was a - - - - - ?

After the above questions were asked, no attempt was made to probe for additional information. The Rorschachs were scored using the notations employed by Beck.

For present purposes verbal fluency was measured by counting the total number of words spoken, number of adjectives, and number of verbs.

The following criteria were used to evaluate the verbal output.

Nouns used as adjectives are counted as adjectives. Examples: *psychology* course; *toy* rifle.

Proper nouns used as adjectives are not counted. Examples: *Baker* Hall; *South* America.

Where colloquialisms or incorrect grammar are used, part of speech designated is that of the subject. Examples: It's real quiet. *Real* here is an adverb and *quiet* is the adjective.

Verbs used as nouns (gerunds) are not counted. Example: I can't get too much studying done. Studying is the gerund.

Dangling participles are counted as adjectives. *Going* into Columbus, High Street runs north and south.

Verbs used as adjectives, participles, are counted as adjectives. Example: You'd probably see people standing in groups. Standing modifies people.

Possessive forms of pronouns are not counted as adjectives, possessive adjectives.

The verb *to be* was not counted as a verb.

Auxiliaries were not counted as separate verbs. These were part of the main verb.

When standing alone they were counted.

Infinitives were not counted.

Limiting adjectives were not counted. Examples: This, those, some, many.

Ah's and similar "feeling" statements were not counted as words.

Not all of the Rorschach scoring categories were used in study. The method used in selecting the scoring categories was that these categories be relatively unaffected by the subject's verbal fluency; for example "affect" and "anxiety." The Rorschach attempts to get at basic personality variables such as "fantasy," "emotionality," "control," etc. From Rorschach literature it was felt that selecting the crucial determinants would not be related to verbal fluency. Although the superficial aspects of personality vary as a function of the situation, the core of the personality remains constant. On this basis, it would be expected that verbal fluency should be unrelated to Rorschach determinants.

Product-moment correlations were obtained among all the variables used at the present time. These variables included Rorschach scores and verbal measures. The resulting correlations were then subjected to a factor analysis using a multiple extraction process to facilitate handling of the data. The criterion of simple structure was employed in the rotation of the factors extracted.

Inspection of the correlational matrix, Table 1, shows, for this sample, disagreement with expectations derived from Rorschach theory. An interesting finding is obtained when M is correlated with *Sum C*. The obtained correlation of 0.54 is difficult to interpret staying within the realm of Rorschach theory; one

Table 1
Intercorrelations of Rorschach Determinants and Verbal Measures

Variable	OSPE	Total Words	Adjectives	Verbs	*M*	*R*	*W%*	Σ*C*	*A*+*Ad*%	*Y*	*V*
OSPE	X	.31	.33	.28	.20	.07	.15	.11	.19	.22	.33
Total Words	.31	X	.91	.97	.40	.51	.09	.51	−.04	.66	.15
Adjectives	.33	.91	X	.82	.33	.48	.05	.33	.08	.57	.15
Verbs	.28	.97	.82	X	.42	.45	.15	.56	−.01	.62	.15
M	.20	.40	.33	.42	X	.41	.14	.54	−.26	.48	.52
R	.07	.51	.48	.45	.41	X	−.29	.67	−.16	.81	.17
W %	.15	.09	.05	.15	.14	−.29	X	.13	.04	−.07	.03
Σ*C*	.11	.51	.33	.56	.54	.67	.13	X	−.39	.79	.30
A + *Ad* %	.19	−.04	.08	−.01	−.26	−.16	.04	−.39	X	−.17	−.26
Y	.22	.66	.57	.62	.48	.81	−.07	.79	−.17	X	.24
V	.33	.15	.15	.15	.52	.17	.03	.30	−.26	.24	X

might expect that the correlation should approximate zero or there should be an inverse correlation for a "normal" population. It seems necessary to conclude for this sample that there is a moderate degree of overlap between the *M* and *C* categories. Another pertinent finding is the correlation of 0.79 between *Y* and *C*. This suggests that these two determinants are not as distinct as claimed in the Rorschach literature. According to Rorschach theory, the correlation should approximate zero; it would seem rather difficult to maintain that *Y* and *C* are measuring different behaviors.

The factor analysis was based on thirty cases. With a sample of this size, the resulting factor loadings should be interpreted with some degree of caution. It is only when the loadings go above approximately .30 that some importance can be attached to them.

Table 2
Factor Matrix

Variable	Verbal Intelligence	Productivity	Elaboration	Individuality
OSPE	67	03	−02	−03
Total Words	50	34	10	06
Adjectives	47	41	02	−03
Verbs	39	32	39	05
M	30	32	12	59
R	10	96	01	00
W %	33	−31	30	09
Σ*C*	15	58	70	33
A + *Ad* %	31	29	−13	−52
Y	35	82	40	−01
V	43	15	−03	58

Four factors were obtained from a factor analysis of the correlational matrix, Table 2. The first factor has its highest loading on the OSPE with the next highest occurring on the verbal measures. The variables *Y*, *V*, *M*, and *W%* also have appreciable loadings on the first factor obtained. These are behavioral indicators which contribute to the loading of this factor but which should not do so according to Rorschach theory. The hypothesis might be ventured that the Rorschach determinants are not indicators of different behavioral manifestations but rather a priori categories which have relatively little independence. Perhaps *Y*, *V*, and *M* reflect the ability to verablize in a testing situation. If the subject cannot express himself verbally, it becomes almost impossible to score shading and vista determinants. Shading responses are supposed to indicate depression; but if the individual is depressed and refuses to verbalize, it is questionable whether or not he will produce *Y* responses. The first factor will be referred to as one of "Verbal Intelligence." Since *Y* per cent and *M* do not have high loadings on this factor, it is questionable that they are adequate indicators of intelligence. This finding is in keeping with other studies [1] which have shown that little relationship, if any, exists between *M* and intelligence.

The second factor extracted is interesting in view of the fact that some Rorschach workers feel that there is relatively high independence of Rorschach determinants; this factor will be referred to as one of "Productivity." Variable *R* has by far the maximum loading which is followed by *Y* and *Sum. C.*

One conclusion appearing warranted for the sample used in this study is that the more responses the subject gives, the greater the possibility of giving some form of color or shading responses. Similar results have been obtained by Wittenborn [9] and by Schroder [8]. Again we see that Rorschach determinants have much in common; such being the case, some of the Rorschach categories might profitably be deleted or combined.

The third factor obtained is referred to as an "Elaboration" factor, that is the extent to which the individual can describe his specific perceptions. Color and shading appear to contribute a maximum to this factor. As with the previous factors we see that the color response is not a distinct category but is related to the other determinants. "Anxiety" and "affect" should not appear correlated for a "normal" population to the extent that they do in this study as well as in that by Wittenborn.

The last factor extracted is called an "Individuality" factor. Movement, vista, and color have high positive loadings while the animal percentage has a high negative loading. This factor agrees somewhat with Rorschach literature in that *M* and *A* are inversely related.

Summary

Thirty undergraduate students were administered individual Rorschachs. The subjects also gave verbal descriptions of five hypothetical situations. Intercorrelations were obtained for some of the Rorschach scoring categories and three verbal descriptive scores. The correlations were then subjected to a factor analysis. The following four factors were obtained: "Verbal Intelligence," "Productivity," "Elaboration," and "Individuality." From the loadings on some of the factors, for the present sample, it appears that Rorschach determinants are related to language measures. Specific Rorschach determinants are sometimes closely related to other determinants, in this matrix, although they previously have been considered relatively independent of the related variables. An analysis of the factors suggests that some revision is called for in the categories of Rorschach scores, and consequently the interpretations of these categories.

Received July 29, 1952.

References

1. Altus, W. D. The group Rorschach as a measure of intelligence at college level. *Amer. Psychologist*, 1949, **4**, 388. (Abstract)
2. Beck, S. J. *Rorschach's test*, Vol. I & II. New York: Grune & Stratton, 1944.
3. Eichler, R. M. Some comments on the controlling of differences in responses on the Rorschach Test. *Psychol. Bull.*, 1951, **48**, 257–259.
4. Hsu, E. H. The Rorschach responses and factor analysis. *J. gen. Psychol.*, 1947, **37**, 129-138.
5. Jessor, R. Personal communication, 1951.
6. Rorschach, H. *Psychodiagnostics.* New York: Grune & Stratton, 1942.
7. Rotter, J. B. The present status of the Rorschach in clinical and experimental procedures. *J. Pers.*, 1948, **16**, 304-311.
8. Schroder, H. Personal communication, 1951.
9. Wittenborn, J. R. A factor analysis of Rorschach scoring categories. *J. consult. Psychol.*, 1950, **14**, 261-267.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

The Effectiveness of Rorschach Elements in the Discrimination Between Neurotic and Ambulatory Schizophrenic Subjects[1,2,3]

Glenn W. Rieman
Veterans Administration Hospital, Altoona, Pa.

Since its introduction in 1921, the Rorschach test [18] has gained widespread use as a technique for the understanding and interpretation of personality. Some clinicians have developed qualitative and quantitative signs to aid in the differential diagnosis between nosological groups. Miale and Harrower-Erickson [11] proposed nine "neurotic signs" to distinguish between neurotic and normal subjects. Rorschach [1, p. 233] contended that certain signs were pathognomonic for schizophrenia. In a preliminary study, Harrower-Erickson [5, p. 114] maintained that her signs might prove to be promising in discriminating between neurotics and psychotics.

The purpose of this study was to determine whether there are statistically significant group differences between neurotic and ambulatory schizophrenic subjects with respect to certain Rorschach elements and signs. The investigation is actually a combination "element-incidence and sign" study since it involved both a preliminary search for group differences in specifically defined elements preparatory to the formulation of signs, as well as the employment of signs previously utilized. The sign approach has been used since it is an objective technique which indicates the salient features in an individual's personality structure with a minimum expenditure of time and effort. This type of investigation holds that the various components of the Rorschach test may be utilized singly or in combination, as contrasted with the intuitive or global approach alone.

Piotrowski [14, 15, 16] described ten Rorschach signs for the identification of patients with intracranial involvements. Twenty schizophrenic signs were gleaned from the literature by Klopfer and Kelley [10]. Buhler, Buhler, and Lefever [2] devised a "Diagnostic Sign List" to aid in the establishment of diagnostic-sign patterns for different clinical groups. Ross [19] applied the sign techniques of Piotrowski and of Miale and Harrower-Erickson. Davidson [4] reported a high correlation between personality patterns of children as classified by an othodox use of the Rorschach test and personal and social adjustment as ascertained by her objective Rorschach signs. Munroe [13] developed her Inspection Rorschach whereby the protocol may be scanned by means of an objective checklist in order to procure a general adjustment rating. To measure personality changes in persons undergoing psychotherapeutic treatment, Muench [12] employed 22 Rorschach factors or patterns of adjustment.

Subjects

Neurotic and ambulatory schizophrenic subjects were chosen for this study because clinicians frequently have difficulty in discriminating between them diagnostically. The neurotic

---

[1] This article is based upon a dissertation submitted by the writer in partial fulfillment of the requirements for the Ph.D. degree, University of Pittsburgh, 1950.

[2] Published with the permission of the Chief Medical Director, Department of Medicine and Surgery, Veterans Administration, who assumes no responsibility for the opinions expressed or conclusions drawn by the author.

[3] The writer is indebted to Dr. Roy M. Hamlin, Chief Clinical Psychologist of the Western Psychiatric Institute and Clinic, University of Pittsburgh, Pittsburgh, Pennsylvania, and formerly Chief Clinical Psychologist, Veterans Administration Mental Hygiene Clinic, Pittsburgh, Pennsylvania, for a critical reading of this paper.

and the schizophrenic groups were divided into two numerically equal subgroups by the odd-even method so that there were four research groups: 50 neurotics and 50 schizophrenics for the original investigation, 50 different neurotics and 50 different schizophrenics for the replication study. Each subject chosen had a service-connected disability for either neurosis or schizophrenia. Furthermore, the diagnosis must have been confirmed currently by a psychiatrist in the installation where the patient was seen for diagnosis, treatment, or both.

It was impossible to determine whether a Rorschach test had been administered prior to the service-connected diagnosis and might have had some influence upon this diagnosis. It is recognized that most Veterans Administration installations do employ the Rorschach test almost routinely and that such results are utilized by many psychiatrists in arriving at a diagnosis. Complete control over this variable was not feasible within the scope of this study. Presumably, a patient would not be placed in either diagnostic category without one diagnosis little influenced or uninfluenced by Rorschach results. In practice, diagnoses continuing over a period of time are based primarily upon clinical behavior rather than upon test results.

Each neurotic subject had been interviewed on an outpatient basis at the Veterans Administration Mental Hygiene Clinic, Pittsburgh, Pennsylvania. Forty-four of the schizophrenic subjects were chosen from the same installation; 42 had been seen at the Veterans Administration Mental Hygiene Clinic, Newark, New Jersey; and 14 were selected from the neuropsychiatric wards of the Veterans Administration general hospitals at Aspinwall, Pennsylvania, and Butler, Pennsylvania.[4] The ambulatory schizophrenic subjects used were individuals who still maintained some contact with reality and had not deteriorated to the extent that constant hospitalization was indicated.

Only male, white, veterans of World War II were used. The schizophrenic subjects from the Pittsburgh and Newark areas were amalgamated in a random fashion; there was no significant difference between the original and the replication studies with respect to the number of Pittsburgh and non-Pittsburgh cases involved. The two nosological groups were equated on the basis of age, intelligence, and education. Ages ranged from 19 through 38. Educational levels varied from the sixth grade through one year of graduate school. Full-Scale Wechsler-Bellevue Intelligence Quotients ranged from 78 through 135. The compared groups were not at all significantly different with respect to age, intelligence, and education.

Procedure

Certain factors, ratios, patterns, and signs, designated as elements, had been compiled previously into a Rorschach Element List[5] composed of 86 items. Many of the elements tested are not signs in the strict sense of the word. A number of them are simply the various location, determinant, and content factors as used by Hertz [6] and Klopfer and Kelley [10]. Some are percentages, ratios, differences, and combinations of factors extracted from sign lists and techniques employed by Muench [12], and Klopfer and Davidson [9]. Some are qualitative characteristics which have been defined quantitatively and checked to indicate their presence. Still others are signs, in the accepted sense, which have been adopted or adapted from devices utilized by Harrower-Erickson [5], Buhler, *et al* [2], and Piotrowski [14, 15, 16].

Each Rorschach protocol was scored by the investigator from the original verbatim copy according to a prescribed procedure which may be described as a modified Klopfer system. Form accuracy, location, and populars were checked against Hertz's Tables [6]. In order to check the accuracy of the scoring, a sampling of the records was rescored independently by another experienced Rorschach worker, using the same scoring guides, after which a percentage of agreement was calculated. Upon completion of the scoring process, the items on the Element List were tabulated for each individual record and these tabulations were amalgamated for the nosological groups under consideration. The 86 Rorschach elements were as follows:

---

[4] Appreciation is extended to the Veterans Administration clinical psychologists who made available their psychological files and records for the accumulation and analysis of data. They are: Dr. David Cohen, Deshon Hospital, Butler, Pennsylvania; Dr. Howard W. Goodman, Aspinwall Hospital, Aspinwall, Pennsylvania; Dr. Roy M. Hamlin, Mental Hygiene Clinic, Pittsburgh, Pennsylvania, and Dr. Samuel B. Kutash, Mental Hygiene Clinic, Newark, New Jersey.

[5] The descriptive definitions of the elements used can be found in the unpublished doctoral dissertation of the writer [17] on file in the University of Pittsburgh library.

I. Scoring Factors

1. W
2. D
3. Dr
4. S
5. Do (oligophrenic)
6. M
7. $M+$
8. FM
9. m (Fm, mF, m)
10. Fk
11. kF
12. k
13. K
14. KF
15. FK
16. $F+$
17. $F-$
18. Fc
19. cF
20. c
21. FC'
22. $C'F$
23. C'
24. FC
25. CF
26. C
27. P (popular)
28. O (original)
29. H
30. Hd
31. A
32. Ad
33. At
34. Sex
35. Ar (artificial)
36. Na (natural)
37. Pathognomonic content
38. Pathological content

II. Responses and Timing

39. R (number of responses)
40. T/R (average time per response)
41. Reaction time to cards without color (I, IV, V, VI, VII)
42. Reaction time to cards with color (II, III, VIII, IX, X)
43. Per cent responses to color (responses to VIII, IX, X)
44. Per cent responses to shading (responses to IV, VI, VII)

III. Form-Level Rating (FLR)

45. Total
46. Average per response
47. Average per response (noncolor)
48. Average per response (color)

IV. Patterns and Ratios

49. $F\%$
50. $F+\%$
51. $FM +$ total $m : Fc +$ total $c +$ total C'
52. $A\ (A + Ad)\ \%$
53. $H + A : Hd + Ad$
54. $W : M$
55. $R+\%$
56. $[(Fk + FK + Fc + FC') - (kF + k + KF + K + cF + c + C'F + C')]$ weighted
57. $[FC - (CF + C)]$ weighted
58. $Fk + FK + Fc + FC'$
59. $F : (Fk + FK + Fc + FC')$
60. $M + FC - (CF + C)$
61. $M - FM$
62. $H - Hd$
63. Sum C (weighted)
64. M : Sum C (weighted)

V. Pathological Signs (Check presence)

65. Perseveration
66. Impotence
67. Perplexity
68. Automatic phrases
69. Color naming
70. Shading shock
71. Color shock
72. Description
73. Significant comments
74. Abstraction and symbolism
75. Confabulation
76. Contamination
77. Sex confusion or unusual handling
78. Position
79. Turning
80. Edging
81. $O-$
82. Self-reference
83. Pathological verbalization

VI. Miscellaneous

84. "W"
85. Rejections or failures (I through X)
86. Succession

Chi square and its probability values were utilized as the statistical tools for a test of significance of the difference between the two groups with respect to each element on the Rorschach Element List. In order to exercise some control over the possible influence of the subject's number of responses (R) upon certain category scores, two precautions were taken. First, the usual method of converting raw scores into percentages by dividing by R was employed where applicable. Secondly, an adaptation of a method advocated by Cronbach [3] was utilized to aid in partialling out differences in R. Each nosological group was divided at the approximate median of the R into two subgroups labeled "Brief" and "Longer" for separate tests of significance. If the p values for these two subgroups differed by less than 0.25 on each element, the "Brief" and "Longer" subgroups were ignored and only the "total" chi square and p value for the element was considered. However, if the p values for these two subgroups differed by 0.25 or more, each "Brief" and "Longer" subgroup was tested separately by chi square for significant differences. In order to obtain a 2×2 table for the application of chi square, each element or element subgroup was dichotomized as near as possible to the median obtained by inspection. Elements scored for presence alone were dichotomized on the basis of presence or absence. Yates's correction for continuity [7] was used with all elements except one. Another correction was made for the few elements the cell frequencies of which were less than five. Because of small cell frequencies, several elements with similar theoretical meanings were combined; a few others could not be treated statistically for the same reason.

Since Rorschach factors are known to vary from one sample to another, further confirmation of the results of the original study was considered worth while. Those elements which revealed differences between groups at the 10 per cent level of confidence or better were singled out for additional study. The 10 per cent level was chosen as the cut-off point to allow for the inclusion and investigation of borderline factors. Those elements were retested on another sample of identical size employing the same techniques used in the original investigation.

Certain limitations of this investigation are recognized quite frankly. A major problem in dealing with nosological groups is diagnosis, especially where there is an overlapping of symptoms as in the case of neurotics and ambulatory schizophrenics. The accuracy and consistency of psychiatric diagnoses have been questioned frequently. In this study, inconsistencies of diagnosis were kept to a minimum by two procedures: (*a*) the selection of subjects was dependent upon two diagnostic judgments per individual with a time lapse between diagnoses; and (*b*) psychiatric diagnosis in both the armed forces and the Veterans Administration employs a well-standardized, tabulated, and similar plan. A second limitation lies in the problem of the homogeneity of the nosological groups under consideration since all neurotic subjects were selected from one mental hygiene clinic whereas the schizophrenics were chosen from three other installations in addition to the one just mentioned. However, both clinics are conducted under similarly oriented psychiatrists. Moreover, since deteriorated schizophrenics are not assigned permanently to the two hospitals, it would appear permissible to include schizophrenic patients from these hospitals under the general classification of ambulatory schizophrenics. A third limitation centers in the applicability of the findings. This study is concerned with groups, and any conclusions derived therefrom are applicable to similar groups alone, in the strictest interpretation. Individuals are the primary concern of clinicians. However, many persons feel that group results may serve as a base of operations for use in individual cases.

Results

With respect to the major scoring categories, the percentages of agreement on the scoring accuracy check are as follows: location, 90; determinants excluding form quality, 87; form quality of determinants, 88; content, 92; populars, 82; and originals, 33.

Table 1

Elements Showing Significant Group Differences on Both Original Investigation and Replication Study

Element	Original investigation chi square	p	Replication study chi square	p	Group exceeding on element specified
Ave. *FLR*	7.844	.002	9.000	.005	Neurotic
Ave. *FLR* (non-color)	4.000	.05	4.910	.03	Neurotic
$R + \%$	7.844	.007	5.762	.02	Neurotic
Perseveration	3.646	.06	5.018	.03	Schizophrenic
Confabulation	6.498	.01	4.158	.04	Schizophrenic
Pathological Verbalization	8.678	.006	8.384	.006	Schizophrenic

According to the established criterion, twenty elements are influenced by the number of *R*. Only 19 of the 79 elements finally tested out in this investigation show significant differences at the 10 per cent level or better between the neurotic and the ambulatory schizophrenic groups. They are: S (Brief); $M\%$ (Longer); $FC'\%$ (Longer); $P\%$ (Total); $H\%$ (Longer); T/R; RT, Color; Average FLR; Average FLR, Noncolor; Average FLR, Color; Succession; $F+\%$; $W:M$ (Longer); $R+\%$; $Fk+FK+Fc+FC'$ (Longer); Perseveration; Shading Shock; Confabulation; and Pathological Verbalization.

The six elements which revealed significant group differences at the 10 per cent level of confidence or better on both the original investigation and the replication study are listed in Table 1.

The possibility that the diagnostic groups were contaminated by the use of the Rorschach in their diagnosis is a justifiable criticism and can be answered finally only by further research. A strong case can be made for the assumption that the clinical use of the Rorschach is not influenced to a major degree by the six items in the specific positive results.

Table 2

Schizophrenic Signs

Average Form-Level Rating	< 0.84
Average Form-Level Rating (noncolor)	< 0.93
$R+$ per cent	< 82
Perseveration	Present
Confabulation	Present
Pathological Verbalization*	Present

*As defined in this study, an element which includes comments and responses that appear queer and peculiar, bizarre, incoherent and vague, illogical and absurd, critical or aggressive, self-depreciatory, irrelevant, morbid to an excessive degree, or neologistic.

While Contamination and Position did not show significant group differences on the original investigation but did so on the replication study, their comparative frequencies suggest that the presence of either type of response is practically pathognomonic for schizophrenia.

The six statistically significant elements tabulated in Table 1 were converted into the "schizophrenic signs" listed in Table 2 by means of cut-off points at the inspectional median in the case of the first three listed, while presence alone was considered a sign in the case of the last three listed. The total number of these signs for each of the individual schizophrenic and neurotic cases used in both studies was determined. These results are presented in Table 3. By means of chi square, it was determined that ambulatory schizophrenics reveal a higher number of these six diagnostic signs, statistically significant at the .001 level of confidence, than do neurotics. If the presence of any three of these six diagnostic signs were chosen arbitrarily as the cut-off point for a differential diagnosis between neurosis and schizophrenia on an individual basis, 65 per cent of the schizophrenics would have been identified properly. However, 29 per cent of the neurotics would have been included falsely within the schizophrenic classification. If five signs were selected as the cut-off point, a mere 1 per cent of the neurotics would have been categorized mistakenly but only 18 per cent of the schizophrenics would have been identified. Such results may have some value for group purposes but probably embody too much inaccuracy for individual differentiation.

Table 3

Distribution of Schizophrenic Signs Shown by Subjects in Each Nosological Group

Total signs per subject	Neurotic	Schizophrenic
6	0	3
5	1	15
4	10	18
3	18	29
2	8	11
1	23	13
0	40	11
	$N=100$	$N=100$

Summary and Conclusions

1. Neurotics and ambulatory schizophrenics as groups do not show significant differences in the great majority of the Rorschach elements treated. Finer tools within the elements themselves might prove more productive.

2. Five elements showed significant group differences at the 5 per cent level of confidence on both samples. They are: Average Form-Level Rating, Average Form-Level Rating (noncolor), $R+$per cent, Confabulation, and Pathological Verbalization. Per-

severation has been retained as significant even though it did not quite meet the above-mentioned standard on one of the two samples.

3. This research indicates the value of a replication study in testing Rorschach factors.

4. A worth-while by-product of this study is the demonstration that two trained individuals working independently with identical scoring guides agree to a large extent in the scoring of Rorschach responses.

5. Several observations may be made with respect to the theoretical meanings of the elements showing significant differences between the two diagnostic groups. The neurotic groups have more capacity to perceive commonly seen concepts and are better able to organize and elaborate accurately than the schizophrenic groups. The schizophrenic groups are more inclined to have difficulty in shifting their thought trends, to be unduly elaborative in an unrealistic manner, and to indulge in bizarre and illogical reasoning.

6. Kendig's "Principle of distortion of reality" [8] as crucial in the diagnosis of schizophrenia has been at least partially substantiated by the present study. These schizophrenic groups do tend to confabulate, to perseverate, to verbalize peculiarly, to perceive in an unusual fashion, and to be inaccurate in organizational and elaborative capacity more frequently than the neurotic groups.

7. Diagnostic "schizophrenic signs" obtained from the statistically significant elements may be useful for group differential purposes but are probably too inaccurate for individual diagnostic discrimination. It may be that these two groups are too ill defined as diagnostic categories to offer hope of a clear-cut differentiation between them by means of Rorschach signs.

8. Although the results of this study are predominantly negative, one observation should be made. Since the two nosological groups employed overlap in symptomatology, results which are only moderately positive or indicative of a trend are actually more significant than are apparently better results with more divergent groups.

9. It is highly improbable that the positive results obtained can be adequately explained by the possibility that the criterion groups may have been contaminated by the utilization of the Rorschach in their diagnosis. However, additional research is needed to verify this contention.

10. The comparative efficiency of this Rorschach sign technique and the global approach in the diagnostic differentiation of these two nosological groups might be worthy of study.

Received June 13, 1952.

References

1. Bochner, Ruth, & Halpern, Florence. *The clinical application of the Rorschach Test.* (2nd Ed.) New York: Grune & Stratton, 1945.
2. Buhler, Charlotte, Buhler, K., & Lefever, W. D. *Development of the basic Rorschach score with manual of directions.* Los Angeles: Rorschach Standardization Studies, No. 1, 1948.
3. Cronbach, L. J. Statistical methods applied to Rorschach scores: a review. *Psychol. Bull.,* 1949, **46**, 393-429.
4. Davidson, Helen H. *Personality and economic background: a study of highly intelligent children.* New York: King's Crown Press, 1943.
5. Harrower-Erickson, Molly R. The value and limitations of the so-called "neurotic signs." *Rorschach Res. Exch.,* 1942, **6**, 109-117.
6. Hertz, Marguerite R. *Frequency tables to be used in scoring responses to the Rorschach ink-blot test* (3rd Ed.) Cleveland: Department of Psychology, Western Reserve Univer., 1946.
7. Kelley, T. L. *Fundamentals of statistics.* Cambridge: Harvard Univer. Press, 1947.
8. Kendig, Isabelle V. Rorschach indications for the diagnosis of schizophrenia. *J. proj. Tech.,* 1949, **13**, 142-149.
9. Klopfer, B., & Davidson, Helen H. *The Rorschach method of personality diagnosis: individual record blank.* New York: Rorschach Institute, 1942.
10. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* (2nd Ed.) Yonkers, N. Y.: World Book Co., 1946.
11. Miale, Florence, & Harrower-Erickson, Molly R. Personality structure in psychoneuroses. *Rorschach Res. Exch.,* 1940, **4**, 71-74.
12. Muench, G. A. An evaluation of nondirective psychotherapy by means of the Rorschach and other indices. *Appl. Psychol. Monogr.,* 1947, No. 13.
13. Munroe, Ruth L. Prediction of the adjustment and academic performance of college students. *Appl. Psychol. Monogr.,* 1945, No. 7.
14. Piotrowski, Z. A. On the Rorschach method and its application in organic disturbances of the central nervous system. *Rorschach Res. Exch.,* 1936-37, **1**, 23-40.
15. Piotrowski, Z. A. The Rorschach ink-blot

method in organic disturbances of the central nervous system. *J. nerv. ment. Dis.*, 1937, **86**, 525-537.

16. Piotrowski, Z. A. Positive and negative Rorschach organic reactions. *Rorschach Res. Exch.*, 1940, **4**, 147-151.
17. Rieman, G. W. An investigation to determine which Rorschach elements reveal group differences between neurotic and ambulatory schizophrenic subjects. Unpublished doctor's dissertation, Univer. of Pittsburgh, 1950.
18. Rorschach, H. *Psychodiagnostics.* (4th Ed.) New York: Grune & Stratton, 1949.
19. Ross W. D. The contribution of the Rorschach method to clinical diagnosis. *J. ment. Sci.*, 1941, **87**, 331-348.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Productivity on the Rorschach as a Function of Order of Presentation[1]

Louis Joseph Maradie
University of Kentucky

Increased productivity in the all-color cards (Figs. VIII, IX, X) is expected on the theory that color has a stimulating effect upon affectively labile individuals. Rorschach himself believed that there was a relationship between the number of responses influenced by color and the extent of emotional excitement although he states, "The causes of this correlation remain to be discovered" [8, p. 98].

A study reported by Dubrovner [4] in which he compared the effects of standard versus achromatic plates revealed no significant shifts in productivity upon retesting. The "color cards" (Figs. VIII-X) produced more responses than the "noncolor cards" (Figs. I-VII) regardless of the presence or absence of color. In view of this it becomes tenuous to assume that color alone is responsible for increased productivity in the last three cards.

In order to provide for controlled comparisons, retesting of the same subjects under different conditions has been deemed necessary. Rabin and Sanderson [7] retested their subjects with both standard and reversed orders of presentation and upon retesting found an increase in stereotypy (*A%*), reduction of productivity in the reverse order condition, and a reduction of "shock" in the retest irrespective of which order was first presented. It appears that with the novelty of the Rorschach situation somewhat eliminated in all but the first acquaintance with the test, comparisons between test and retest data are incongruous.

Schachtel [9] believes that both reaction time and productivity are affected by the testee's fears, attitudes, and interests which are activated by the testing situation. It seems that memory factors would strongly influence the subject's definition of the retest situation for the challenge of the now familiar and no longer enigmatic plates is dispelled.

A common clinical observation is that the initial interview situation tends to be less productive than later ones. The subject finds himself in a novel relationship which, with the establishment of rapport, becomes decreasingly disturbing to him as a frame of reference is gradually established. In the unstructured Rorschach situation, the phenomenon is doubly significant. In empirically determining the order of the plates within the series, Rorschach predicted that failure on the first plate would rarely be encountered [8, p. 16]. He realized that early-appearing hesitancy and evasiveness could be alleviated by a relatively simple introductory stimulus card. Initial restraint and anticipatory anxiety would disappear as adaptation to the task developed.

In using the latin square, this present study is designed to circumvent the artifacts which arise by retesting the same subjects under different conditions. Its specific purpose is to investigate the possibility that irrespective of the sequential positions of the cards, later-appearing cards will produce more responses than earlier cards. At the same time, this design permits an evaluation of the productivity attributable to the cards themselves, independent of their order of appearance.

---

[1] This study was done in partial fulfillment of requirements for an M.S. degree in psychology. The writer wishes to express his sincere appreciation for the indispensable guidance and suggestions of Dr. Richard M. Griffith and Dr. Robert D. North, his counselors and friends. Grateful acknowledgment is also made to the student nurses of Good Samaritan Hospital, Lexington, Kentucky, who, in providing the data for this experiment, graciously gave of their time.

Procedure

From a female nurses training school, a group of 22 volunteers were obtained. Their ages ranged from 17 to 28 years with a mean chronological age of 20.2 years. Of the 22 subjects tested, 2 were omitted because of extreme overproductivity of autistic quality. In such a study dealing with a limited number of subjects, variant scores offer a definite procedural problem. Dangers of bias in omitting data must be balanced with dangers of including discrepant data which would prevent later generalization to other defined and homogeneous groups. Rorschach authorities have indicated that the response totals for normal subjects rarely exceed fifty [1, 6, 8]. The response totals for the two subjects excluded were 115 and 143 respectively — both considerably more than twice as large as the next highest score of 49 responses.

In the latin-square design as used here, systematically randomized standard Rorschach plates were presented to each subject [5, p. 302]. Ten distinct sequential orders were arranged in which each card preceded and followed every other card just once [3]. This arrangement may be made more apparent by examination of Table 1 in which each card appears once in each column and once in each row. In this square, any systematic influence resulting from the positional or sequential effect of the color cards would thus tend to be minimized.

To each of these discrete orders, a subject was randomly assigned and presented with the plates peculiar to that order. The same latin square was replicated by the random assignment of an additional subject to each of the ten orders of presentation. In this manner, twenty different subjects responded to ten dissimilar orders of presentation.

The instructions to the subjects were those suggested by Beck [2, p. 2]. After having given two or more associations for a figure, any declaration by the subject that she was through with it was accepted. Encouragement took place when less than two responses per figure were obtained but was progressively reduced until it was entirely discontinued after the fifth card.

Results

Before the data from the two separate latin squares were combined, they were tested for homogeneity of total variance by means of Bartlett's test [5, p. 195]. The resulting chi-square value of .683 was not significant and offered no evidence against the hypothesis of random sampling from a common population.

In comparing the data of the separate latin squares, homogeneity of variance was also established for the residual mean square and the sequential position mean square. The *F* values obtained were 1.04 and 1.16 respectively, which were not significant at the 5 per cent level.

After combining the data from the separate latin squares it was further ascertained that differences in productivity between the sequences or pairs of subjects responding to one particular order in the latin square were not significantly different. Examination of the agreement between the samples as shown in

Table 1

Latin-Square Design of the Order of Presentation of the Rorschach Plates

	Positions									
Orders	1	2	3	4	5	6	7	8	9	10
1	I	II	III	IV	V	VI	VII	VIII	IX	X
2	II	IV	VI	VIII	X	I	III	V	VII	IX
3	III	VI	IX	I	IV	VII	X	II	V	VIII
4	IV	VIII	I	V	IX	II	VI	X	III	VII
5	V	X	IV	IX	III	VIII	II	VII	I	VI
6	VI	I	VII	II	VIII	III	IX	IV	X	V
7	VII	III	X	VI	II	IX	V	I	VIII	IV
8	VIII	V	II	X	VII	IV	I	IX	VI	III
9	IX	VII	V	III	I	X	VIII	VI	IV	II
10	X	IX	VIII	VII	VI	V	IV	III	II	I

Table 2

Sums of Responses for Orders of Sequential Presentation (Note Table 1)

Order of Presentation	Sample 1	Sample 2	Total
1	34	43	77
2	18	29	47
3	31	40	71
4	33	22	55
5	21	23	44
6	24	26	50
7	49	23	72
8	25	27	52
9	15	23	38
10	21	37	58

Table 3

Analysis of Variance Results

	S. S.	*df*	M. S.	*F*
Independent observations:				
Order of presentation	76.32	9	8.48	1.20 (not sig.)
Residual between individuals (error)	70.60	10	7.06	. . .
Total between individuals	146.92	19		
Correlated observations:				
Cards	35.32	9	3.92	4.45**
Positions	43.52	9	4.84	5.48**
Residual from latin square	78.36	72	1.09	1.23 (not sig.)
Residual within individuals	79.40	90	.88	. . .
Total within individuals	236.60	180		
Total	383.52	199		

**Significant at 1 per cent level.

Table 2 and the insignificant *F* value of 1.20 in Table 3 demonstrate this fact.

Table 4 reveals that the productiveness of these 20 subjects is apparently greater in the later sequential positions. The mean number of responses in the first seven positions was 51.1 while that of the last three positions was 68.7. A *t* value calculated for the difference between these means was 4.00, which was significant at the 1 per cent level. Testing the significance of the mean square for total responses by positions against the residual mean square within individuals, it was found that the obtained value of *F* (5.48) was highly significant for 9 and 90 degrees of freedom (Table 3).

Table 4

Aggregate Sums of Responses by Positions

Positions	Responses	Positions	Responses
1	52	6	48
2	46	7	59
3	51	8	63
4	48	9	66
5	54	10	77

The other factor of interest was the effect of the cards themselves. When tested for significance against the residual mean square, the mean square for cards was also found to be highly significant with an *F* value of 4.45 (Table 3).

Further analysis of the effect of color upon productivity demonstrated that when the number of responses of the color cards (Fig. II, III, VIII, IX, X) was compared to those of the completely achromatic cards (Fig. I, IV, V, VI, VII), the difference between the means ($t = 1.80$) was not significant. When the number of responses to Figures I-VII was compared to those to figures VIII-X, the difference between the means was significant ($t = 2.12$) at the 5 per cent level. Examination of Table 5 shows that Card X, in contrast to Cards VIII and IX, is by far the most provocative of the series. Since Cards VIII and IX are also completely chromatic and did not elicit any noticeable increment in productivity, it appears that the configuration of Card X may be the crucial factor facilitating the subjects' responsiveness.

Table 5

Aggregate Sums of Responses by Cards in All Possible Positions

Cards	Responses	Cards	Responses
I	47	VI	57
II	55	VII	58
III	55	VIII	53
IV	53	IX	59
V	48	X	79

Summary and Conclusions

The data used in this experiment were obtained from 20 of 22 Rorschach-naive student nurses. The 2 subjects were discarded because of their extreme overproductivity. The 20 subjects were presented with ten distinct systematically randomized sequential orders of standard Rorschach plates in a latin-square design. In this manner each plate appeared in all positions and preceded and followed every other plate an equal number of times. Within the limitations imposed by the homogeneity of the sample used, examination and statistical analysis of the data revealed that:

1. Irrespective of the order of the cards, the *position* of the cards is of importance with later-appearing cards producing more responses than earlier-appearing cards.

2. Irrespective of their orders of presentation or their sequential relationships, the cards themselves differ in the number of responses elicited with Card X, because of its color or configuration, or both, evoking more responses than any other.

Thus, it is shown that the positional effect, quite apart from color, could account for increased productivity in later-appearing cards. While made suspect, it is not definitely shown that color does not also have some effect. The present writer expects to isolate this factor by the presentation of achromatic plates in the same latin-square design. If similar results are obtained in this follow-up study, it will provide conclusive evidence that the configuration of the plates and the order of their presentation are the factors responsible for increased productivity in the later-appearing cards.

Received July 30, 1953.

References

1. Beck, S. J. *Rorschach's test.* Vol. II. New York: Grune & Stratton, 1949.
2. Beck, S. J. *Rorschach's test.* Vol. I. New York: Grune & Stratton, 1950.
3. Bugelski, B. R. A note on Grant's discussion of the latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.,* 1949, **46**, 49-50.
4. Dubrovner, R. J. The effect of color on productivity and reaction time in the Rorschach test. *J. clin. Psychol.,* 1950, **6**, 331-336.
5. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1951.
6. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1942.
7. Rabin, A., & Sanderson, M. An experimental inquiry into some Rorschach procedures. *J. clin. Psychol.,* 1947, **3**, 216-225.
8. Rorschach, H. *Psychodiagnostics.* Berne: Hans Huber, 1942.
9. Schachtel, E. G. Subjective definitions of the Rorschach test situation and their effect on test performance. *Psychiatry,* 1945, **8**, 419-448.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Rorschach Characteristics of Persons Scoring High and Low in Acceptance of Self[1]

Robert E. Bills
University of Kentucky

A technique called "An Index of Adjustment and Values" which is designed to give measures of the concept of self, acceptance of self, the concept of the ideal self, and maladjustment has been reported by Bills, Vance, and McLean [2]. Additional studies by Roberts [6] and Bills [3] have confirmed further the reliability and validity of the Index scores.

The present study is concerned with one of the four measures of the Index—acceptance of self. Studies by Seeman [8], Snyder [10], Sheerer [9], and Stock [11], have established that during successful psychotherapy there is an increase in positive attitude toward self or acceptance of self. Since acceptance of self has been considered a central personality variable, it follows that people who are high in acceptance of self should differ, in general, in their personality characteristics from people who are low in acceptance of self.

The Rorschach test [1, 4, 7] purports to give a description of an individual's personality. If persons who are high in acceptance of self differ in personality characteristics from people who are low in acceptance of self, and if the Rorschach is able to differentiate personality characteristics, then persons who are high in acceptance of self should have Rorschach test results which differentiate them as a group from persons who are low in acceptance of self.

Design

Twenty volunteers, female students at the University of Kentucky, were tested with the Index of Adjustment and Values and the Rorschach test. This pilot group was divided into two groups on the basis of acceptance-of-self scores above and below the mean score of the group on which the Index was standardized. The division placed 15 subjects who had scores below the mean of acceptance of self in one group and 5 subjects who had scores above the mean of acceptance of self in the other group. A comparison of the Rorschach results for the two groups suggested the hypotheses listed in Table 1.

A second group of 50 volunteer students was obtained by using as the criterion of selection, scores on acceptance of self as measured by the Index of Adjustment and Values. Twenty-five of the subjects had scores which were in excess of one standard deviation above the mean acceptance-of-self score of the standardization population of the Index (this group will be referred to as the "high scorers") and twenty-five of the subjects ("low scorers") had scores more than one standard deviation below the mean acceptance-of-self score of the standardization population. The two groups were balanced for sex and Kentucky General Classification Test scores. Each group consisted of 8 male and 17 female subjects, and a test of the significance of the difference between the means of the scores of the two groups on the Kentucky General Classification Test gave a t of .750 which showed that at the .60 level the groups were not significantly different in general ability. The groups were not significantly different in respect to the number of responses on the Rorschach, as may be seen in Table 1.

Each of the 50 experimental subjects was tested with the Rorschach, using Beck's instructions [1], by an examiner who had no knowledge of the hypotheses under investiga-

---

[1] This study was made possible by a grant from the University of Kentucky Research Fund Committee. The author is deeply appreciative for this aid. The author would also like to express his thanks to Glen E. Roberts who worked as an assistant on the project.

Table 1

The Hypotheses of the Study and the Statistical Results

Variable	Hypothesis	Chi square	*p*
M:Sum C	Above Sum $C > M$ Below $M \geq$ Sum C	14.346	.001
$F+\%$	Above $F+\%$ lower than below	35.280	.001
FC	Above less FC	9.680	.01
CF	Above more CF	27.000	.001
$CF:FC$	Above $CF \geq FC$ Below $FC > CF$	38.782	.001
Sequence	Above fewer cards in sequence		n.s.
$A\%$	No difference in two groups		n.s.
$F\%$	No difference		n.s.
R	No difference		n.s.
FC'*	No difference	2.080	.20
M	Above less M	2.000	.20
Maps, etc. (evasion)	Above more evasive	2.225	.20
$A+H$: $Ad+Hd$	Above $<2\frac{1}{2}$ whole to 1 part. Below $>2\frac{1}{2}$ whole to 1 part.	3.125	.08
P	Above more P	4.023	.05
$\%R$ VIII–X	% smaller in above group	8.013	.01
RT Chromatic	No difference		n.s.
RT Achromatic	No difference		n.s.
RT Dif. Chr. and Ach.	Above greater difference	2.922	.10
$W:M$	Above $W > 2M$ Below $W \leq 2M$	4.667	.05
T/R	Above react quicker	3.945	.05
No. Turns	Above fewer turns	2.885	.10
No. Cards Turned	Above turn fewer cards	2.000	.20
$M:FM$	$FM \geq M$ in above $FM < M$ in below	8.323	.01
Defense†	Above more defensive	2.885	.10
$H\%$	No difference		n.s.
$c:k$‡	No difference		n.s.
$W\%$	Above larger $W\%$		n.s.
Obj.	No difference		n.s.
N	No difference		n.s.
Fire,etc. (aggression)	No difference		n.s.
(H)	Below more (H)		n.s.
Fm‖	Above more Fm		n.s.

*Includes FC', C', and $C'F$.

†Includes Fc, c, cF, k, kF, Fk, K, KF, and FK.

‡Ratio of $Fc + c + cF$ to $k + kF + Fk + K + KF + FK$.

‖Includes Fm, m, and mF.

tion. The Rorschachs were then scored by a second person, not the Rorschach examiner, who had no knowledge of the subjects' classification on the Index of Adjustment and Values. The scoring system used was a mixture of Beck [1], Klopfer and Kelley [4], and Krugman [5]. Beck's system was used to score populars and location, and to determine form level. In the determination of $F+\%$ Beck's tables were used to determine form level, and $F+\%$ was calculated from the formula: Good F divided by Rorschach F. Rorschach F is the total number of determinants minus the color and human movement responses. All other scoring was from Klopfer and Kelley except as noted at the bottom of Table 1.

The records of two of the "high scorers" and two of the "low scorers" were selected at random and scored by two judges working independently of each other.[2] Neither of the judges was acquainted with the basis of selection of the subjects nor the hypotheses being tested. Percentages of agreement between the Rorschach scorer and the two judges were calculated for each of the four cases for the location, determinant, content, and "popular" categories. The percentages ranged from 76 to 100, and 23 of the 32 percentages were 90 per cent or greater.

Results

Table 1 presents the Rorschach variables investigated, the predicted direction of the differences of the high scorers and low scorers, and the chi squares and probabilities resulting from the calculations from double dichotomous tabulations. In these tabulations one dichotomy was formed by the independent variable, acceptance of self, and the other dichotomy by the division indicated in the prediction, or where no definite prediction was made concerning the dividing point, at the median score of all 50 cases.

Discussion

The data show striking differences in the two experimental groups. Only two of the low scorers had $F+\%$'s less than the median of all 50 cases, and only two of the high scorers had $F+\%$'s greater than the median. It may be concluded, therefore, that the high scorers and low scorers are distinctly different in respect to $F+\%$.

---

[2] The author is appreciative of the assistance given by O. S. McLean and R. C. Nichols, who served as the judges.

The number of human movement responses was not significantly different in the two groups, although the difference approached statistical significance. A comparison of the number of human movement responses with the weighted color sum, however, shows another striking difference—the experience balances of the two groups are reversed. Twenty-two of the high scorers showed the sum of the weighted color responses greater than the number of M and one case had M equal to the sum of the weighted color responses, whereas, 16 of the 25 low scorers had the number of M greater than the sum of the weighted color responses. This finding leads to the conclusion that the high scorers are extratensive in their experience balance while the low scorers are introversive.

The high scorers and low scorers differed significantly in the type and number of color responses. The low scorers used fewer color responses but more "well-controlled" color represented by FC, whereas the high scorers used a greater number of color responses and more "poorly controlled" color represented by a predominance of CF and C responses. No "pure C" appeared in the records of the low scorers, but 14 of the 25 high scorers showed at least one pure C response. Furthermore, the high scorers showed an excess of CF over FC, and the low scorers an excess of FC over CF. Of the 25 high scorers, 24 had more CF than FC responses, but 23 of the low scorers had more FC than CF. It is concluded, therefore, that if the type and number of color responses is representative of affective life, then the high scorers have a stronger and less well-controlled affect than the low scorers.

Other variables including number of populars, the percentage of responses on the last three cards, the $W:M$ ratio, the time per response, and the $M:FM$ ratio show that at the .05 level of significance, or better, that the two groups are clearly different.

It is apparent, therefore, that in respect to certain important personality characteristics the high and low scorers, as measured by the Rorschach, constitute two distinctly different personality groups. Thus, it is concluded that the Index of Adjustment and Values is able to separate groups with different personality characteristics.

Summary

On the basis of a pilot study, 34 specific hypotheses were formulated relative to Rorschach variables present in the personalities of persons high and low in acceptance of self as revealed by the Index of Adjustment and Values. Fifty new subjects with extreme acceptance of self scores were examined with the Rorschach test. Of the 34 hypotheses 23 were confirmed, 4 were denied, and 7 remain in doubt.

It was concluded that distinct Rorschach personality characteristics distinguish subjects who are high in acceptance of self from those who are low in acceptance of self.

Received July 19, 1952.

References

1. Beck, S. J. *Rorschach's test*: I. New York: Grune & Stratton, 1949.
2. Bills, R. E., Vance, E. L., & McLean, O. S. An index of adjustment and values. *J. consult. Psychol.*, 1951, **15**, 257-261.
3. Bills, R. E. Progress of research with the index of adjustment and values. Mimeographed paper, Univer. of Kentucky, Lexington, 1952.
4. Klopfer, B., & Kelley, D. M. *The Rorschach technique*. Yonkers: World Book Co., 1946.
5. Krugman, M. Class lectures, Teachers Coll., Columbia Univer., Spring, 1948.
6. Roberts, G. E. A study of the validity of the index of adjustment and values. *J. consult. Psychol.*, 1952, **16**, 302-304.
7. Rorschach, H. *Psychodiagnostics*. Berne, Switzerland: Hans Huber, 1942.
8. Seeman, J. A study of the process of nondirective therapy. *J. consult. Psychol.*, 1949, **13**, 157–168.
9. Sheerer, Elizabeth T. An analysis of the relationship between acceptance of and respect for self and acceptance of and respect for others in ten counseling cases. *J. consult. Psychol.*, 1949, **13**, 169–175.
10. Snyder, W. U. An investigation of the nature of nondirective psychotherapy. *J. gen. Psychol.*, 1945, **33**, 193-223.
11. Stock, Dorothy. An investigation into the interrelations between the self concept and feelings directed toward other persons and groups. *J. consult. Psychol.*, 1949, **13**, 176–180.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Adjustment and the Discrepancy Between Self Concept and Inferred Self

A. D. Calvin and Wayne H. Holtzman
The University of Texas

One of the most widely accepted beliefs among psychologists today is that discrepancy between the self concept and objective reality is a common feature of maladjustment. Such discrepancies are said to represent a lack of insight, and a major goal of psychotherapy is to help the individual reorganize his self concept so that it will correspond more adequately with the self as viewed by others. Granted that in general the above statement may be true, it does not follow, without further experimental verification, that the relationship between such discrepancies and severity of maladjustment is as simple and direct as is often assumed. The present study was undertaken to examine the nature of this relationship in several groups of college men.

Description of the Study

Population. The population consisted of 79 male students, all the members of four fraternities, A, B, C, and D, who were residing in their respective fraternity houses at the time. Each subject was well acquainted with all the others in his group, having lived continuously in the chapter house for at least five months. The number of men in fraternities A, B, C, and D was 23, 17, 20, and 19, respectively. Groups of this size are small enough to allow extensive individual interaction and yet sufficiently large to permit adequate statistical analysis.

Collection of the data. All of the subjects were first given the Minnesota Multiphasic Personality Inventory (MMPI) in group sessions. The standard instructions for administering the test were followed [3]. At the time of testing an appointment was made to see each subject individually in order to gather data concerning his self-evaluations and his judgments of the other members in his group.

At the beginning of the individual session, each subject was interviewed briefly concerning his personal history in order to establish rapport and gain certain background information. Following the interview the subject was presented with a set of small cards, each containing the name of one of his associates and a code number for recording purposes. He was then asked to rank all the members of his group, including himself, on seven personality traits. Having the individual rank himself as well as the rest of his group made possible the study of discrepancies between the self- and group-judgments of each individual. A description of each personality trait was typed on a plain, white card to facilitate the ranking. The traits used, in the order of presentation to the subject were:

1. *Leadership.* Place the individual whom you think would be most desirable as a leader to represent the group in campus affairs at the *top*; rank the others in order beneath.
2. *Tolerance.* The general disposition of an individual to be fair and patient toward people whose attitudes are opposed to his own.
3. *Adjustment.* Well adjusted, handles personal problems well, rarely shows signs of excessive worry or anxiety, seems basically happy.
4. *Drive.* Drives himself onward, always tries hard, very ambitious.
5. *Tactfulness.* Tactful, aware of the feelings of others, keen, shows good discretion, subtle, not blunt, polished and well-mannered.
6. *Gregariousness.* Very gregarious, the "life-of-the-party" type, mixes well and seems to enjoy being around people.
7. *Social Understanding.* The needs, motivations, interest, and activities of the individuals in a group are numerous. Molded together, these diverse needs, etc., give the group a common (held-by-the-majority) set of goals, objectives, and pur-

poses. At the *top*, place the individual whom you think has the best understanding of—and works hardest to achieve—these common goals, objectives, and purposes.

These particular traits were chosen because they had been used successfully in previous studies [5, 6] employing similar ranking methods.

The interview and ranking session was conducted in complete privacy and each subject was told that his judgments would be guarded in strict confidence from the rest of the group. The subjects were informed that opportunities to discuss their own MMPI scores would be available at the conclusion of the study, an arrangement which served to keep their motivational level high throughout the study.

Method of analysis. Pooled rank scores for each individual on every trait were computed separately for each group by summing the individual ranks assigned him by his associates and placing subjects in rank order according to their pooled scores. An individual's final rank on a given trait thus represented the consensus of opinion of his associates, providing an appropriate measure of one aspect of the *inferred self,* the individual "as an informed other person sees him" [4, p. 381]. Similarly, the rank which the subject assigned to himself on that same trait provided a corresponding measure of one aspect of the *self concept,* "those parts of the phenomenal field which the individual had differentiated as definite and fairly stable characteristics of himself" [8, p. 112].

The trait description for the self-enhancing or more desirable extreme of each trait continuum had been deliberately chosen for use in this study, so that self rank which was higher than the group rank for a particular individual would be indicative of a self-enhancing discrepancy. On the other hand, a self rank which was lower than the group rank would indicate a self-depreciative discrepancy.

At first glance it would appear that a simple discrepancy score could be obtained merely by subtracting the self rank from the group rank. There are two reasons why such a method is unsatisfactory. First, rank scores have a rectangular distribution while the distribution of the underlying trait is more likely to be roughly normal. A discrepancy of, say, three ranks between self and group scores at either extreme of the rank order will represent quite a different psychological distance on the underlying trait continuum than the same numerical discrepancy in the middle range. Second, the direction and size of such a discrepancy score is partly a function of the group rank. For example, an individual who is judged by his associates to be the best adjusted person in his group can only receive a self-depreciative discrepancy score while the most maladjusted individual can only get a self-enhancing score. To the extent that the group rank and discrepancy score are correlated, any conclusion based upon an analysis of such simple discrepancy scores would be an artifact.

The first of these difficulties was largely overcome by assuming that the underlying trait distribution was normal.[1] Tables were constructed for each group for converting ranks into scale values on the baseline of the normal distribution curve so that each of the resultant distributions had a mean of 5.0 and a standard deviation of 2.0 [2, p. 248]. A discrepancy score, $(G-S)$, was computed for every individual on each trait by subtracting his converted self-score, S, from his converted group score, G.

To eliminate the spurious relationship between G and $(G-S)$, the effect of G was partialled out in the following manner: A scattergram was constructed for each of the seven traits by plotting G against $(G-S)$, using all 79 individuals. Since in every case visual inspection of the scattergram showed clearly that the regression line for predicting $(G-S)$ from G was linear with homoscedastic arrays, product-moment correlation coefficients and regression equations for predicting $(G-S)$ from G were computed. As might be anticipated from the earlier discussion, the correlations (presented in Table 1) were sufficiently high in every case to require adjustment of the self-group discrepancy scores. Regression equations were plotted on each scattergram and corrected discrepancy scores were determined by the following equation:

---

[1] The validity of this assumption in the present instance is borne out by the fact that invariably the subject would claim that he had considerable difficulty ranking individuals in the middle ranges, but that as he moved toward either extreme the task became much easier, indicating that the underlying psychological continuum was bell shaped, probably normal.

$$(G-S)_c = (G-S) - (G-S)^1$$

where

$(G-S)_c$ = corrected discrepancy score
$(G-S)$ = original discrepancy score
$(G-S)^1$ = discrepancy score predicted from regression of $(G-S)$ on G.

Table 1

Correlation Between G and $(G-S)$ for Each of the Seven Traits

Trait	*r*
Leadership	.71
Tolerance	.68
Adjustment	.79
Drive	.59
Tactfulness	.73
Gregariousness	.63
Social Understanding	.68

The corrected discrepancy score, $(G-S)_c$, is that part of the original discrepancy between self- and group-judgment for a particular trait and individual which remains after the spurious relationship between G and $(G-S)$ has been eliminated. A positive $(G-S)_c$ can be thought of as indicating a self-enhancing tendency greater than the average individual with a comparable G score, the larger $(G-S)_c$ the stronger the relative amount of self-enhancement. Similarly, a negative $(G-S)_c$ can be interpreted as indicative of less than average self-enhancement or even self-depreciation, as the case may be.

Four measures of the discrepancy between self concept and inferred self were derived for comparison with adjustment as measured by the MMPI. The first consisted of $(G-S)_c$ for the trait *Adjustment*. The second was a more general measure of self-enhancing tendency obtained by summing algebraically $(G-S)_c$ across the seven traits for each individual. This will be referred to as *Sum* $(G-S)_c$. The third and fourth scores, $|(G-S)_c|$ for *Adjustment* and *Sum* $|(G-S)_c|$, were similar to the first two with the exception that the direction of the discrepancy was ignored in each case. In a sense they are measures of *insight* defined simply as the degree of agreement between the inferred self and the self concept.

Results

Before attempting to study the relationships between variables derived from the group- and self-judgments and scores on the MMPI, the reliabilities of these new variables should be examined. The reliability of G, the pooled group score assigned each individual, has been reported by Greenwood [1] in an unpublished study using these same data. Reliability coefficients were computed separately for each of the four groups of subjects by dividing each group into two halves and correlating the pooled scores. As estimated by the Spearman-Brown formula the reliabilities of the final pooled scores varied from .85 for *Tolerance* to .94 for *Leadership*, highly satisfactory for the present purposes.

Although the present data permit no evaluation of the reliability of S, the self-judgment, evidence is available from an earlier study on a similar group of subjects that the reliability of self-judgments is also quite high. From data provided by Ogdon [6] in a longitudinal study of 21 men in a fraternity at the University of Texas, the present authors obtained high test-retest reliability for self-judgments of leadership. The correlation between scaled ranks over a three-week period was .97. After six weeks during which the fraternity elected new officers, calling for sharp changes in the self concepts of several individuals, the test-retest reliability was .90. This suggests that self-judgments obtained by the present methods have a high degree of reliability over a time span of at least several weeks.

Since both G and S are presumably highly reliable, it follows immediately that $(G-S)_c$ and $|(G-S)_c|$ also have sufficient reliability to warrant serious investigation of their relationships to independent measures of personality.

To have clear meaning as a derived variable, *Sum* $(G-S)_c$ should reflect a consistent tendency for an individual to enhance or depreciate himself, relatively speaking, on a variety of personality characteristics. The degree of such self-consistency was determined by correlating the sums of $(G-S)_c$ for traits 1, 2, and 7 with the sums of $(G-S)_c$ for traits 3, 5, and 6. The choice of these particular traits was based upon a cluster analysis of the trait intercorrelations [1] and represents an attempt to make the two sets of scores as comparable as possible. The situation is somewhat analogous to that frequently encountered in test theory where over-all test reliability is estimated from the correlation between two similar halves, although in the present case the halves are only roughly parallel at best. The correlation between the two sets of scores was only .35, a low value but one which is significantly different from zero. A crude estimate of the reliability of Sum $(G-S)_c$ obtained by the Spearman-Brown formula is .56. The low degree of consistency in the discrepancies between self- and group-judgments from one trait to another may be due in large part to considerable variability in the extent to which possession of a given personality characteristic provides self-esteem.

In a similar manner, the split-half correlation for *Sum* $|(G-S)_c|$ was determined and found to be only .19, too low to justify further analysis of this variable. Moreover, the distribution of *Sum* $|(G-S)_c|$, is somewhat U-shaped due to the summing of $(G-S)_c$ without regard for algebraic sign, making comparison with other variables by correlational methods difficult to interpret.

Table 2

Correlations of MMPI Scales with Adjustment and Two Measures of Self-Enhancing Tendency

MMPI Scale	*G* for *Adjustment*	$(G-S)_c$ for *Adjustment*	Sum$(G-S)_c$
Hs	.11	−.17	−.02
D	.29*	−.42**	−.17
Hy	.11	−.03	−.10
Pd	.18	−.16	−.13
Mf	.33*	−.14	.02
Pa	.09	−.16	.03
Pt	.34*	−.44**	−.14
Sc	.30*	−.36*	−.13
Ma	.14	−.03	.11

**r* significantly different from zero at .01 level
***r* significantly different from zero at .001 level

After first adjusting raw scores on the MMPI for *K* and then converting to *T* scores using tables given in the MMPI manual, product-moment correlation coefficients were computed between each of the standard MMPI scales and each of three scores from the rank data: (*a*) $(G-S)_c$ for *Adjustment*, (*b*) *Sum* $(G-S)_c$, and (*c*) *G* for *Adjustment*. The group score for *Adjustment* was included because of interest in the relationship between adjustment as measured by a personality inventory and adjustment as judged by an individual's close associates. Table 2 summarizes the results of this analysis.

The pattern of relationships between scores on the MMPI and *G* is somewhat similar to that for $(G-S)_c$ in spite of the fact that *G* and $(G-S)_c$ have zero correlation as a result of the procedures used to derive $(G-S)_c$. *Sum* $(G-S)_c$, on the other hand, is not significantly correlated with any of the MMPI scales. This is probably due primarily to the unusually low reliability of *Sum* $(G-S)_c$ as a measure of self-enhancing tendency.

To the extent that *D*, *Pt*, and *Sc* are measures of maladjustment, the highly significant relationship between the MMPI and $(G-S)_c$ for *Adjustment* indicates that tendency to enhance the self is inversely related to maladjustment, the more poorly adjusted the individual, the more self-depreciative, relatively speaking, he appears. This finding is in agreement with results obtained by Sheerer working in an entirely different context. In a study of counseling interviews she reported "... a marked and fairly regular increase in the measured acceptance of and respect for self from the beginning to the end of the cases" [7, p. 175]. In each of ten cases the occurrence of self-depreciative statements decreased as the client's adjustment improved during the course of counseling.

The meaning of the relationship between *G* and the MMPI becomes clearer when the structure of the MMPI is examined more closely. In a factor analysis of 15 MMPI scales, Tyler obtained five factors, the first of which was defined chiefly by *D*, *Pt*, and *Sc*, the three scales significantly related to *G* and $(G-S)_c$ in the present study. In interpreting this factor, Tyler states, "It appears . . . to reflect a chronic state of marginal maladjustment including concern with one's self. Subjects who score high on the factor would tend to have difficulties in their social contacts, possibly by reason of their dissatisfactions, worry and disgruntlement, and their complaining and pessimistic attitude" [9, p. 454]. The significant relationship between G and the *Mf* Scale suggests that manifestation of a feminine interest pattern in a fraternity member is sometimes viewed by his associates as a sign of maladjustment.

Using $|(G-S)_c|$ as a measure of insight with respect to adjustment two groups of subjects were formed, a *High Insight Group* consisting of 24 individuals with $|(G-S)_c|$ less than or equal to 0.5, and a *Low Insight Group* comprised of 14 men with $|(G-S)_c|$ greater than or equal to 2.0. Mean T scores for these two groups on each of the MMPI scales are presented in Table 3, together with t values reflecting the significance of the group differences.

Table 3

Comparison of High Insight Group and Low Insight Group on the MMPI

MMPI Scale	Mean T Score Low Gp.	High Gp.	t	p
Hs	56.1	49.2	2.32	.05
D	55.6	54.6	0.22	—
Hy	58.0	55.7	1.13	—
Pd	60.2	52.0	2.72	.01
Mf	66.5	60.7	2.30	.05
Pa	55.6	52.8	1.15	—
Pt	60.2	57.1	0.84	—
Sc	61.0	56.0	1.48	—
Ma	59.3	56.6	0.78	—

The mean of the *Low Insight Group* is consistently higher than that of the *High Insight Group,* sufficiently so for scores in three of the MMPI scales to prove significantly different. This result indicates that individuals who manifest poor insight regarding their own level of adjustment are more likely to be maladjusted than are those who show good insight, substantiating the claim which is frequently made that lack of insight is a sign of maladjustment.

The highly significant difference between the two groups on the *Pd* Scale is of particular interest. Although the MMPI scale designations must not be taken too literally, the frequent interpretation of a high *Pd* score as indicating an inability to profit from experience, a tendency to disregard social mores, and a general lack of insight [3] is entirely consistent with the results of this study.

Summary and Conclusions

All resident members of four fraternities were administered the Minnesota Multiphasic Personality Inventory. In later individual sessions each individual ranked all the members of his fraternity, including himself, on seven personality traits, *Leadership, Tolerance, Adjustment, Drive, Tactfulness, Gregariousness,* and *Social Understanding.* A normalized pooled rank score (G) for each individual on every trait provided an appropriate measure of one aspect of the *inferred self.* The respective self-rank (S) provided a corresponding measure of one aspect of the *self concept.* Adjusted self-group discrepancy scores, $(G-S)_c$, were used to derive four measures of the discrepancy between self concept and inferred self: (a) $(G-S)_c$ for *Adjustment,* (b) *Sum* $(G-S)_c$, (c) $|(G-S)_c|$ for *Adjustment,* and (d) *Sum* $|(G-S)_c|$. The first two served as measures of self-enhancing tendency; the latter two as measures of insight.

The major findings of this study are as follows:

1. Corrected self-group discrepancy scores obtained by the present procedures are sufficiently stable to warrant serious consideration as new measures of personality.

2. Tendency to enhance the self is inversely related to maladjustment, the more poorly adjusted the individual, the more self-depreciative, relatively speaking, he appears.

3. Maladjustment as judged within a restricted normal range by an individual's associates is directly related to maladjustment as measured by a typical personality inventory, the MMPI.

4. Individuals who manifest poor insight regarding their own level of adjustment are more likely to be maladjusted than are those who show good insight.

Received July 24, 1952.

References

1. Greenwood, C. M. Leadership in normal functioning groups. Unpublished master's thesis, Univer. of Texas, 1951.
2. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
3. Hathaway, S. R., & McKinley, J. C. *The Minnesota Multiphasic Personality Inventory.* New York: Psychological Corp., 1943.
4. Hilgard, E. R. Human motives and the concept of self. *Amer. Psychologist,* 1949, **4**, 374-382.
5. Holtzman, W. H. Validation studies of the Rorschach test: Shyness and gregariousness in the normal superior adult. *J. clin. Psychol.,* 1950, **6**, 343-347.
6. Ogdon, D. P. The study and prediction of attitudinal changes in a normally functioning group. Unpublished master's thesis, Univer. of Texas, 1950.
7. Sheerer, Elizabeth T. An analysis of the relationship between acceptance of and respect for self and acceptance of and respect for others in ten counseling cases. *J. consult. Psychol.,* 1949, **13**, 169–175.
8. Snygg, D., & Combs, A. W. *Individual behavior.* New York: Harper, 1949.
9. Tyler, F. T. A factorial analysis of fifteen MMPI scales. *J. consult. Psychol.,* 1951, **15**, 451-456.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Recent Research with Personality Inventories

Albert Ellis
New York, N. Y.

Although it is often assumed that projective techniques of personality evaluation have virtually monopolized the field of clinical psychology in recent years, and that paper and pencil tests of personality are rarely employed by reputable clinicians, the fact remains that far more research and theoretical papers on personality questionnaires have been published during the past six years than have been published during any equivalent period in the history of psychology. Indeed, with the possible exception of the Rorschach test, it would appear that more research and clinical use is now being made of personality inventories than of any other type of assessment instrument; and if educational, vocational, counseling, and military applications are included, then it would appear that more children and adults are now being given personality questionnaires than are being given any other type of diagnostic procedure or instrument.

Several comprehensive reviews of the literature on personality inventories have appeared during the past several years, including those by Anderson and Embree [1], Ellis [3, 6, 9], Ellis and Conrad [10], Eysenck [11], Strang [14, 15], Strang and Pansegrouw [16], Traxler [17, 18], and Traxler and Jacobs [19]. The present paper will not be another such review, but will briefly report the results of a study of the effectiveness of personality inventories when they are employed for various kinds of individual and group discriminations. In the course of this study, all research reports on personality questionnaires which appeared in the literature between January, 1946 and December, 1951 were surveyed. The results of these research reports were tabulated to determine whether, according to the criterion and level of significance stipulated by the author of each report, the inventory employed in each study did or did not significantly discriminate among the individuals or groups tested.

A survey of the recent studies of personality inventories reveals the following findings:

1. In 21 studies where intercorrelations were calculated among the subscales of different inventories to determine whether they were significantly related to each other, 20 studies showed significantly high correlations, indicating that the inventories were not measuring as many independent traits as their manuals claimed that they were.

2. In 26 studies where scores on one personality inventory were correlated with scores on another inventory purporting to measure equivalent traits, 12 correlations were found to be significantly high while 14 were not.

3. In 12 studies where inventory scores were correlated with Rorschach test results, significant relationships were found in 2 instances but were not found in the other 10 studies. Significant relationships between inventory scores and other projective technique results were found in 1 instance and not found in 3 other studies.

4. In 25 studies where the ability of subjects to fake personality inventory scores was tested, it was found that they could and did fake significantly better or worse scores in 22 instances and that they did not do so in 3 instances. In several of the studies demonstrating that personality inventory scores could be faked, it was found that special detection scales, such as the F *minus* K scale of the Minnesota Multiphasic Personality Inventory, could partially spot and compensate for the faking.

5. Results of using personality inventories for making various types of group discriminations are summarized in Table 1. In general, the data in Table 1 would indicate, that the inventories significantly differentiate most oft-

Table 1

Significant Discriminations Found in Research Studies with Personality Inventories, 1946 to 1951

Discriminations Tested	Number of Studies in which Significant Discriminations Were Found	Number of Studies in which no Significant Discriminations Were Found	Total Number of Studies
Inventory scores vs. diagnostic examinations:			
Neuropsychiatric conditions	58	26	84
Psychosomatic conditions	11	2	13
Delinquent, criminal, and psychopathic behavior	15	13	28
Alcoholism	6	1	7
Predicting success in treatment	24	10	34
Total	114	52	166
Inventory scores vs. personality ratings, by associates, instructors, or counselors:			
Personality ratings	8	8	16
Sociometric ratings	5	5	10
Total	13	13	26
Inventory scores vs. behavioral characteristics:			
Extent of social participation	6	3	9
Leadership success	4	1	5
Propensity to remain healthy	5	5	10
Success in courtship and marriage	3	2	5
Tendency to become a sexual deviant	2	1	3
Tendency to have a favorable self concept	5	2	7
Extrasensory perception ability	2	2	4
Success at motor tasks	7	3	10
Tendency to take pronounced attitudes on controversial questions	4	3	7
Miscellaneous behavioral characteristics	15	19	34
Total	53	41	94
Inventory scores vs. vocational and academic test and performance results:			
Tendency to have specific vocational interests	11	5	16
Membership in different vocational groups	12	7	19
Success in actual vocational performance	18	23	41
Success in academic achievement	4	21	25
Total	45	56	101
Inventory scores vs. group differences:			
Intellectual groups	12	14	26
Educational groups	2	2	4
Age groups	6	1	7
Sex Groups	13	4	17
Socioeconomic groups	3	7	10
Ethnic groups	5	0	5
Marriage and family groups	5	5	10
Disabled and ill groups	3	7	10
College groups	6	1	7
Regional and national groups	2	2	4
Religious groups	2	2	4
Veterans groups	1	1	2
Miscellaneous groups	5	1	6
Total	65	47	112
Grand Total	290	209	499

en when used with psychodiagnostic groups and least often when used with vocational and academic groups, with their discriminative efficiency when used with other kinds of criteria or groupings falling somewhere in between.

At first blush, it is difficult to say whether the results of these recent researches with personality inventories are disappointing or not. On the one hand, they would seem to indicate that the inventories give reasonably successful discriminations when used with neuropsychiatric, psychosomatic, alcoholic, age, sex, ethnic and college groups. On the other hand, they would seem to indicate that the inventories are not measuring the independent traits they are supposed to measure; that they do not agree too well with each other nor with the results of Rorschach and projective tests; that they are easily faked; and that they discriminate unsuccessfully when used with vocational, academic, socioeconomic, and disabled and ill groups. Above all, these results would indicate that, in virtually none of the areas in which they are commonly used, do personality questionnaires *consistently* show significant group discriminations. On the contrary, different experimenters keep obtaining directly contradictory results when using the same or similar inventories.

Actually, the results shown by this survey are less favorable for personality inventories than cursory examination would indicate. Detailed analysis of many of the studies surveyed shows the same methodological failings that Ellis and Conrad [10] pointed out in a previous review. That is to say, many of the studies in which significant discriminations were obtained are suspect because they involved criterion contamination or overlap, inadequate or overly diffuse samples, set-up experiments, differential motivation on the part of the subjects, and statistical and other shortcomings. Moreover, since personality inventories are normally designed for individual diagnosis or assessment, since most of the surveyed studies are of group discriminations, and since these group results are rarely consistently discriminative, it is clear that the validity of these instruments for the main specific purpose for which most of them have been designed is, as yet, far from being solidly established.

Perhaps the main failing in many of the studies surveyed is the tendency of many authors to assume that, merely because significant *score* differences have or have not been found, significant *trait* differences have or have not been proved. One researcher after another keeps concluding that "members of the unaccepted group are more likely to have personality disturbances than members of the accepted group" [2]; or "a significant relationship was found to exist between one's personality adjustment and his occupational interests" [12]; or the like. Such conclusions, of course, will not be warranted until it is conclusively shown, beyond a shadow of a doubt, that personality scores *do* have a consistently high relationship with "personality disturbance," "personality adjustment," "neuroticism," or other traits which their authors and publishers claim they have.

The results of the present study, as well as of previous surveys made by the writer [4, 5, 6, 7, 8, 9, 10], would lead to the conclusion that although modern personality inventories, when employed for group diagnosis, have *some* degree of validity, scores obtained by individual subjects are partly or largely dependent upon (*a*) the subject's specific motivation at the time of taking the test; (*b*) the conditions under which he is tested; (*c*) his intelligence; (*d*) his degree of psychological sophistication; (*e*) his general test-taking attitudes; (*f*) his general psychological defenses; (*g*) his specific (conscious and unconscious) tendencies to lie to others and to himself; and (*h*) various other motivational and attitudinal factors. To consider these factors adequately, a good many "correction" scales or devices must be included in modern inventories—such as the various correction scales which have been and still are being added to the MMPI. So far, however, such correction scales have been shown to have only a limited effectiveness; and it may be that some of the most needed of these scales—e.g., one which would indicate the extent to which subjects unconsciously defend themselves against answering inventory questions in an objective manner—are almost impossible to devise, validate, and effectively incorporate into testing procedure because of the intrinsic limitations of personality questionnaires.

When suitable correction and qualitative analytic procedures are devised for use with inventories, they normally add enormously to the scoring and interpretation time of the tests. Thus, utilizing a test like the MMPI today involves scoring the correction as well as the clinical scales, interpreting all the scales globally in terms of the correction factors and in relation to all the other scales, and doing a profile analysis rather than simple score interpretations.

The question therefore arises: Assuming that the newer scoring and interpretation techniques give a more valid picture of the subject's "personality" than do the older methods (an assumption which has not as yet been experimentally validated), are they worth the time and effort involved? After all, personality inventories were originally devised not because no other or better methods of assessing certain human "traits" were known during the First World War (when the Woodworth Personal Data Sheet became the first of a long line of such instruments), but because they presumably offered a *quicker* and *more convenient* method of "personality" assessment than did conventional psychiatric interviews and other diagnostic techniques. If, now, as is evident from the material in Hathaway and Meehl's monumental MMPI *Atlas* [13], the average clinical psychologist must take considerable time and trouble to master the technique of scoring and interpreting inventory results, and then must spend an hour or two scoring, graphing, and interpreting the results of each individual test given, the vaunted "advantages" of using inventories become most dubious.

Otherwise stated: since personality inventories depend, in the last analysis, on printed questions, and since virtually no one claims for them the advantage of getting at unconscious and semiconscious material which effective clinical use of interview and projective methods are generally conceded to some extent to uncover, it is difficult to see why clinicians should spend considerable time first mastering and then using these inventories. The clinical psychologist who cannot, in the time it now takes a trained worker to administer, score, and interpret a test like the MMPI according to the best recommendations of its authors, get much more pertinent, incisive, and depth-centered "personality" material from a straightforward interview technique would hardly appear to be worth his salt.

Summary

A study was made of all researches on personality inventories published between January, 1946 and December, 1951. It was found that in most cases inventory scores discriminate significantly when used with neuropsychiatric, psychosomatic, alcoholic, age, sex, ethnic, and college groups. On the other hand, it was found that in most instances the inventories are not measuring the independent traits they are supposed to be measuring; they do not agree too well with each other nor with the results of Rorschach and projective tests; they are easily faked; and they usually do not give significant group discriminations when used with vocational, academic, socioeconomic, and disabled and ill groups. It was especially found that, in none of the areas in which they are commonly employed, do personality inventories consistently show significant group discriminations. It was pointed out that when these inventories are most effectively used, they tend to become equally as time consuming as alternative psychodiagnostic procedures, such as interviewing and the use of projective techniques, that seem to be more clinically incisive and valuable.

Received June 9, 1952.

References

1. Anderson, G. V., & Embree, R. B. Appraisal of the individual. *Rev. educ. Res.*, 1948, **18**, 157-174.
2. Brody, D. S. A genetic study of sociality patterns of college women. *Educ. psychol. Measmt*, 1951, **10**, 513–520.
3. Ellis, A. The validity of personality questionnaires. *Psychol. Bull.*, 1946, **43**, 385-440.
4. Ellis, A. A comparison of the use of direct and indirect phrasing in personality questionnaires. *Psychol. Monogr.*, 1947, **61**, No. 3 (Whole No. 284).
5. Ellis, A. Discussion of Heinlein's comment on "The validity of personality questionnaires." *Psychol. Bull.*, 1947, **44**, 83-86.
6. Ellis, A. Personality questionnaires. *Rev. educ. Res.*, 1947, **17**, 53-63.
7. Ellis, A. The value of marriage prediction tests. *Amer. sociol. Rev.*, 1948, **13**, 710-718.
8. Ellis, A. The relationship between personality inventory scores and other psychological test results. *J. soc. Psychol.*, 1948, **28**, 287-289.

9. Ellis, A. Self-appraisal methods. In L. E Abt, & D. Brower (Eds.), *Progress in clinical psychology*. New York: Grune & Stratton, 1952.
10. Ellis, A., & Conrad, H. S. The validity of personality inventories in military practice. *Psychol. Bull.,* 1948, **45**, 385-426.
11. Eysenck, H. J. Personality tests: 1944-1949. In G. W. T. H. Fleming, *Recent progress in psychiatry*. London: Churchill, 1950. Vol. II., pp. 197-253.
12. Feather, D. B. The relation of personality maladjustments of 503 University of Michigan students to their occupational interests. *J. soc. Psychol.,* 1950, **32**, 71-78.
13. Hathaway, S. R., & Meehl, P. E. *An atlas for the clinical use of the Minnesota Multiphasic Personality Inventory*. Minneapolis: Univer. of Minnesota Press, 1951.
14. Strang, Ruth. Methods, technics, and instruments of mental hygiene diagnosis and therapy. *Rev. educ. Res.,* 1946, **16**, 436-443.
15. Strang, Ruth. Technics and instruments for the improvement of mental health. *Rev. educ. Res.,* 1949, **19**, 386-394.
16. Strang, Ruth, & Pansegrouw, Deborah. Studies of individuals. *Rev. educ. Res.,* 1948, **18**, 382-395.
17. Traxler, A. E. Measurement in the field of personality. *Education,* 1946, **66**, 424-430.
18. Traxler, A. E. Nonprojective technics in the appraisal of personality. *Growing points in educational research.* Washington: American Educational Research Association, 1949. Pp. 324-330.
19. Traxler, A. E., & Jacobs, R. Construction and educational significance of structured inventories in personality measurement. *Rev. educ. Res.,* 1950, **20**, 38-50.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Distortion in Paranoid Schizophrenia[1,2]

Paul G. Daston, Gerald F. King
VA Hospital, Ft. Custer, Michigan, and Michigan State College

and Stewart G. Armitage
VA Hospital, Ft. Custer, Michigan

On the basis of clinical observations and experimental evidence schizophrenic thought processes have been described as being rigid, concrete, and distorted. However, experimental studies in this area have not provided consistent support for most of these generalizations. The purpose of this research was to investigate distortion in schizophrenia using a controlled quantifiable approach.

In reference to schizophrenic behavior, what might be called distortion has been described in many ways, ranging from Beck's [3] "inaccurate perception" of clearly implied forms to Cameron's [4] "displacement" of precise verbal terms by approximate concepts. In Aborn's review of the literature, he concluded that although "it may be regarded as secondary to some more fundamental notion of abnormality, distortion is in its broadest sense primarily descriptive of the introduction of contradictory elements into a situation providing no external barriers to consistent behavior" [1, p. 7].

Methodology

The methodology employed in this research stems from the work of Asch [2], Goodman and Lorge [5], and Aborn [1]. Using descriptions of hypothetical people, Asch was interested in how an over-all impression of a person was formed. From the responses of his subjects to a checklist of trait names, he found that an impression of personality is a function of an interaction of traits rather than of simple summation. A variation of this procedure was developed by Goodman and Lorge and utilized in a study of the dynamics involved in the formation of personality impressions by seventh- and ninth-grade school children. Aborn modified the technique as it is used in the present research.

Studying distortion in schizophrenia, the latter investigator constructed a number of short stories describing hypothetical people. Some of the stories utilized positive trait adjectives only, and some contained only negative traits. Further, one set of descriptive stories explicitly stated trait names, while another simply implied them. For example, the all-positive story with trait names explicitly stated was as follows:

> The first things people notice about him are that he is cheerful and friendly. In school, he always got good marks and was looked upon as being intelligent. He is very active and takes part in all sports. As a player, he is known to be fearless and is respected because he is honest.

In comparison, the all-positive story with the traits implied was:

> On his way home, he found a pocketbook in the street and turned it over to the policeman on the corner. He smiled and greeted most of the people he passed along the street and stopped to talk with the man at the newspaper stand. The newspaper man likes to listen to him because he usually knows what is going on in the world. Sometimes they talk about baseball, since they both play on the neighborhood team. They have known each other ever since he chased and caught two men who had robbed the newspaper stand.

---

[1] This research was presented in part at the meetings of the Midwestern Psychological Association in Cleveland, April, 1952.

[2] Published with permission of the Chief Medical Director, Department of Medicine and Surgery, Veterans Administration, who assumes no responsibility for the opinions expressed or the conclusions drawn by the authors.

The authors are especially indebted to the cooperation and assistance provided by Dr. Murray Aborn of Michigan State College.

Immediately after reading each story, the subject indicated his "picture" of the person in the story by checking appropriate words on a word-selection checklist, which consisted of 36 positive and 36 negative adjectives. The checklist was constructed to sample all areas of personality description, with the affective tone of the adjectives being based on the ratings of 62 judges. For each story, the subject's responses could be quantified in terms of *return of givens, expansions,* and *distortions.* A distortion was operationally defined as a response which was inconsistent with the given trait pattern of a story; for example, checking a negative adjective for a positive story.[3]

While Aborn's results failed to show any statistically significant differences between a mixed group of schizophrenics and normals, a trend toward greater distortion on the part of the experimental group was noted. It seemed possible that this trend might be accentuated if more opportunity were allowed for the schizophrenics to introject the stimulus material by inserting a time interval between the presentation of the stories and the checklist.

Experimental Procedure and Subjects

In testing this hypothesis, the following two modifications in methodology were employed: (*a*) the experimental group was confined to paranoid schizophrenics; and (*b*) a temporal factor was introduced. After reading each story, the subject spent five minutes on a simple motor task in which he was required to cross out *O*'s and circle *X*'s. Then he was presented with the checklist and asked to check all the words he would use to describe the person in the story. This procedure was followed for all four stories. The stories were presented in a different order to different subjects to minimize any transfer effects that might be involved.

The paranoid schizophrenic and normal groups both contained 27 male subjects. The groups were equated for age, education, and intelligence, as shown in Table 1. The subjects were selected from a rather limited age range; it actually extended from about 22 to 35 years. The measure of intelligence was a short form of the Wechsler-Bellevue Scale, Form 1 (consisting of the Comprehension, Arithmetic, and Similarities subtests), which Rabin [6] and others have shown correlates very highly with the full scale with both normals and neuropsychiatric patients. In order to reduce the possibility that some of the words in the stories or checklist might be misunderstood by the subjects and thus contaminate the results, the minimum IQ established for the selection of subjects was 100. Further, since none of the words in the stories and checklist were, according to the Thorndike-Lorge word count [7], above seventh-grade comprehension (an educational level below all the subjects), this was deemed an adequate safeguard. An attempt was also made to select the two groups so there would be an equivalent range of occupations.

Table 1

Group Characteristics of the Paranoid Schizophrenics ($N = 27$) and the Normals ($N = 27$)

Group	Age		Education		Intelligence IQ	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Normals	29.6	4.0	12.3	2.5	115.8	10.0
Schizophrenics	30.6	3.9	12.0	2.6	116.1	12.2

The subjects of the experimental group had all been given a final diagnosis of "schizophrenia, paranoid type" by a staff of hospital psychiatrists. The majority were open-ward patients, and all were cooperative and in good contact. All the normals were gainfully employed, had never been hospitalized for neuropsychiatric reasons, and were not currently undergoing treatment.

Results

A comparison of the two groups on the total number of adjectives used for each story, summarized in Table 2, reveals no differences. None of the computed *t* ratios are above 1.00. The data were then analyzed on the basis of return of givens for each story, which refers to the checking of trait adjectives stated directly or strongly implied in the stories. As can be seen in Table 3, there was a tendency for the

---

[3] For a more detailed account of Aborn's methodology, a microfilm of his complete manuscript can be obtained from University Microfilms, Ann Arbor, Mich. Publ. No. 2336, Pp. 84.

Table 2

Comparison of the Experimental and Control Groups on the Total Number of Adjectives Used

Group	Total Number of Adjectives Checked Stories* 1–A Mean	1–A *SD*	2–A Mean	2–A *SD*	1–B Mean	1–B *SD*	2–B Mean	2–B *SD*
Normals	24.52	8.69	17.78	8.07	24.59	9.37	18.11	6.70
Schizophrenics	26.81	8.56	20.37	10.41	23.70	8.85	19.37	8.01

*Key: 1, all positive traits; 2, all negative traits; A, traits stated directly; B, traits implied.

Table 3

Comparison of the Experimental and Control Groups on Return of the Givens

Group	Return of Givens Stories 1–A Mean	1–A *SD*	2–A Mean	2–A *SD*	1–B Mean	1–B *SD*	2–B Mean	2–B *SD*
Normals	5.33	0.94	4.63	1.39	5.11	1.02	4.41	1.49
Schizophrenics	5.00	1.19	4.11	1.81	4.70	1.24	3.78	1.90
t ratio	1.14		1.18		1.29		1.33	

normals to return more givens and the schizophrenics to be more variable. Neither trend, however, reaches statistical significance.[4]

In analyzing the distortion scores, the chi square yielded by Bartlett's test of homogeneity showed that homogeneity of variances for the two groups could not be assumed, which introduces certain difficulties for standard statistical treatment. An inspection of the data suggested that the extreme scores of six schizophrenic subjects (high in distortion) accounted for the marked variance of the experimental group. When the extreme distortion scores made by these subjects (a total of 11 scores) were eliminated and the chi square was computed again, homogeneity of variances could then be accepted. We now felt justified in employing analysis of variance since discarding these extreme scores depressed the mean distortion values for the experimental group, thus working against our hypothesis. As such, it would constitute a conservative or "weak" test of significance.

The results of the analysis of variance are presented in Table 4. It can be seen that the paranoid schizophrenics produced significantly more distortion than the normal controls (beyond the 1 per cent level of confidence). Comparing the groups on individual stories showed significant differences in stories 2-A (1 per cent level) and 1-B (5 per cent level) which does not suggest any particular pattern. The interaction term between the diagnostic categories and stories was not significant, indicating the experimental subjects did not react to the materials in a different manner than did the controls. It was a reaction of greater magnitude but in the same direction.

Table 4

Treatment of Distortion Scores by Analysis of Variance

Source of Variation	*df*	Sum of Squares	Mean Square	*F*	*p*
Total	204	282.71			
Diagnostic Categories	1	13.74	13.74	19.35	<.01
Stories	3	17.81	5.94	8.37	<.01
Individuals	53	144.87	2.73	3.85	<.01
Stories vs. DC	3	4.56	1.52	2.14	NS
Error	144	101.73	0.71		

In summary, the paranoid schizophrenics showed both significantly more distortion and more variability than did the normals. It should be pointed out, however, that this difference in amount of distortion is a group difference, with considerable overlapping between groups. As such, it is not adequate for individual prediction.

---

[4] These results coincide with the findings of Aborn who used return of givens as one index of rigidity.

Discussion

In light of Aborn's findings, it would appear that schizophrenics do not necessarily deviate perceptually from normals in the immediate recall of controlled verbal material. However, when a temporal interval is interpolated between presentation of the material and recall, distortion occurs. This is seemingly a result of the interaction of a more or less correct perception with the intervening schizophrenic personality variables. The results suggest that it is the processes underlying schizophrenic thought or judgment that somehow go awry.

It is realized that there are certain difficulties in comparing the findings of this research with those of Aborn. Since two modifications in his methodology were used, the question arises as to which variable accounts for the distortion. The interpolated time interval has already been discussed. The other possibility lies in the nature of the subjects used. Although Aborn's group of schizophrenics of various subtypes was comparable to that of this research in terms of such characteristics as age, education, and intelligence, our experimental group was confined to paranoid schizophrenics. Since the principle criterion in formulating the diagnosis of "paranoid schizophrenia" is the presence of "projective" thinking (as seen in delusions), this factor may be the crucial one in producing distortion. The answer to this question seems obtainable in further research with this technique in which the following comparisons would be made: (*a*) paranoid schizophrenics vs. normals without the temporal interval; and (*b*) nonparanoid schizophrenics vs. normals with the temporal interval.

Summary

The purpose of this research was to investigate distortion, a characteristic often attributed to the behavior of schizophrenics. The methodology employed four short stories describing hypothetical people and a word-selection checklist. After reading each story, the subject spent five minutes on a simple motor task. He was then presented with the checklist and asked to check all the words he would use to describe the person in the story. A distortion was defined as the checking of a word which was inconsistent with the given trait pattern of a story. The subjects consisted of 27 paranoid schizophrenics and 27 normals, the two groups being equated for age, education, and intelligence. A statistical analysis of the data revealed that the experimental group showed both significantly more distortion and more variability than the normals. The results were compared and discussed in terms of the findings of Aborn, who developed the basic methodology used in this research.

Received June 23, 1952.

References

1. Aborn, M. A study of rigidity and distortion in normals and schizophrenics with controlled verbal material. Unpublished doctor's dissertation, Columbia Univer., 1950.
2. Asch, S. E. Forming impressions of personality. *J. abnorm. soc. Psychol.*, 1946, **41**, 258-290.
3. Beck, S. J. Personality structure in schizophrenia. *Nerv. ment. Dis. Monogr.*, 1938, No. 63.
4. Cameron, N. Reasoning, regression and communication in schizophrenics. *Psychol. Monogr.*, 1938, **50**, No. 1 (Whole No. 221).
5. Goodman, S. Forming impressions of persons from verbal report. Unpublished doctor's dissertation, Columbia Univer., 1950.
6. Rabin, A. I. A short form of the Wechsler-Bellevue test. *J. appl. Psychol.*, 1943, **27**, 320–324.
7. Thorndike, E. L., & Lorge, I. *The teachers word book of 30,000 words.* New York: Bur. Publ., Teachers Coll., Columbia Univer.. 1944.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Some Effects of Sodium Amytal on "Deteriorated" Schizophrenics[1]

Willard A. Mainord
State College of Washington[2]

Conflict still exists as to the nature of schizophrenia, much of the controversy centering upon whether this pathology is of an organic, functional, or psychosomatic nature. Schizophrenic "deterioration," in particular, remains obscure as to nature and genesis, and it is obvious that more knowledge is necessary for best clinical results. Three types of theories are advanced to explain schizophrenic deterioration, i.e., schizophrenia is a progressive, organic disease; schizophrenia is a strictly functional disorder which serves some defensive function in the psychological economy of the individual; and, schizophrenia is a psychosomatic malady, and causal factors as well as disease processes may be both organic and functional. Cameron [2] notes that a rigid organic point of view usually implies actual brain deterioration not amenable to anything but physiological treatment, and with little hope of full restitution of intellectual efficiency.

The usual definition of schizophrenic deterioration is descriptive, but recently Levy and Kennard [7] have advanced a concept of deterioration with more objective criteria. Their study shows a relationship between clinical judgment of deterioration, abnormality of EEG records, and significantly large Wechsler-Bellevue deterioration ratios. These criteria can be duplicated by other investigators, and comprise the operational definition of schizophrenic deterioration used in this paper.

If schizophrenic deterioration can be alleviated without organic change, psychological therapies would appear to offer hope for rehabilitation of deteriorated patients. Layman [5, 6] has demonstrated that it is possible to obtain a temporary improvement in the intellectual performance of an unselected group of schizophrenics through injections of sodium amytal. There is evidence [1, 8, 9] which indicates that these results may come from changes in motivational states which come about as a result of the action of this drug. The problem investigated in this paper can be stated as follows: can a statistically significant improvement in performance be obtained from deteriorated schizophrenics, as defined by the criteria suggested by Levy and Kennard, while the patients are under the influence of sodium amytal? The working hypothesis was that such an improvement could be obtained because of motivational changes resulting from the drug.

The nature of the criteria suggested the following secondary hypotheses: if schizophrenia is a progressively disabling disease, there should be a significant negative correlation between the subjects' ages and their obtained IQ's; and there should be a significant positive correlation between the improvement observed under amytal and the size of the deterioration ratio. This latter hypothesis is based on the assumption that the Wechsler-Bellevue deterioration ratio does measure organic damage, and that the greater the amount of tissue pathology, the greater the amount of potential improvement.

---

[1] This paper is derived from a study done as a master's thesis under the direction of Dr. J. H. Elder to whom grateful acknowledgment is made. Acknowledgment is also made to Dr. Sol Levy, Clinical Director, Eastern State Hospital, Medical Lake, Washington and to the clinical staff for the interest and cooperation shown relating to this study. All conclusions, of course, are the author's own, and the responsibility for them entirely his.

[2] Now at the University of Washington.

Procedure

The criterion given the most weight in the selection of subjects was clinical judgment, inasmuch as behavior itself is basic to any concept of deterioration. As an additional precaution, the theorized progressive nature of the disease was allowed for by selecting only subjects who were considered to have had the illness for at least five years. The abnormality of the EEG records was determined by a competent specialist. The minimum Wechsler-Bellevue deterioration ratio acceptable for the experimental group was 15 per cent after correction for aging—as Wechsler [11] suggests 10 per cent as a minimum significant figure. All subjects who had been subjected to any kind of therapy after the original testing were excluded. Table 1 summarizes the mean age, known duration of illness, deterioration ratio, and IQ for both the experimental and test-retest groups. The test-retest group was used only to give some indication of usual test-retest effects on long-standing schizophrenics, inasmuch as the experimental control was accomplished by comparing pre-amytal with post-amytal scores. All cases were selected from the same population as that used by Levy and Kennard.

Table 1

Mean Age, Known Duration of Illness, Deterioration Ratio, and IQ for Experimental and Test-Retest Groups.

Group	N	Mean Age	Mean Duration	Mean Deterioration ratio*	Mean IQ*
Experimental	22	43.5	13.6	31.4%	84.5
Test-retest	18	41.0	10.4	12.6%	100.6

*The utilization of the criteria for selection so diminished the original population that it was impossible to match these two groups for IQ and for deterioration ratio.

The subjects, experimental and test-retest, were originally tested with the Verbal section of Form I of the Wechsler-Bellevue Intelligence Scale in a two-week period from December 20, 1950 to January 5, 1951. The experimental group was retested approximately nine months later while under the influence of sodium amytal.[3] The final test was administered to the experimental group approximately one week later, again under conventional conditions. The test-retest group was also retested at this time.

Results

The results obtained can be said to bear out the hypothesis that a significant, temporary improvement in measured intellectual performance would be obtained through the administration of sodium amytal.

Table 2

Mean Group Scores of Experimental and Test-Retest Groups on Wechsler-Bellevue Verbal Scale under all Test Conditions

Group	Pre-Amytal	Under Amytal	Post-Amytal
Experimental	84.5	94.1	86.4
Test-retest	100.6	—	101.6

Table 2 summarizes the results, in which it can be seen that the mean Verbal IQ of the experimental group increased to 94.1 from 84.5 under sodium amytal, and dropped back to 86.4 in the post-amytal testing. In other words, the experimental group, while under the influence of the drug, increased its mean Verbal IQ by 9.6 points (significant at better than the .001 level of confidence), and then lost 7.7. points of this gain (significant at better than the .001 level of confidence) when the same test was administered again without the drug. The difference between the means of the pre- and post-amytal conditions was not statistically significant, despite practice effects. The test-retest group improved its performance from 100.6 to 101.6, this one-point gain being comparable to the 1.9-point gain between the first and third testing of the experimental group.

The secondary hypotheses were not confirmed by the data, inasmuch as the correlations obtained in each case were insignificant. Thus, only the primary hypothesis, i.e., that a significant improvement in performance of deteriorated schizophrenics could be obtained by injection of sodium amytal, was supported by the results obtained.

---

[3] The drug was administered intravenously in a solution composed of one grain per cc. of distilled water. No set dosage was used; the drug was discontinued when the subject seemed to have reached a testable level which appeared optimal from observer judgment.

Discussion

If sodium amytal does not temporarily mitigate organic pathology, the obtained data would seem to support the point of view that schizophrenic deterioration is not entirely physiological. In view of other work [8, 9], it would seem logical to ascribe the improvement noted to changes in the patients' motivation. Therapeutically, it might be expected that schizophrenia and its accompanying deterioration may be partly mitigated by psychological methods. A permanent improvement in all spheres of behavior of comparable magnitude to the temporary intellectual improvement obtained in this study would be enough to return many deteriorated schizophrenics to society. Failure of such methods to date may be due to lack of perseverance with difficult patients, inadequate techniques, and lack of funds at the insistence of economy-minded legislators and administrators.

This study suggests that the severity of the schizophrenic process may have little or nothing to do with length of illness. A familiar sight in any institution is the well-preserved paranoid schizophrenic of many years standing. This study further suggests that no matter what Wechsler's deterioration ratio does measure [3, 4, 10, 11], it does not seem applicable to schizophrenic deterioration, as no relationship between these ratios and the data could be detected.[4]

Of striking interest, but difficult to evaluate, were the results obtained through the use of amytal with patients otherwise untestable. Five such patients (none of them otherwise used in this study) were tested, and four of them yielded scorable protocols with Verbal IQ's ranging from 60 to 95, apparently indicating that much potential is still available to many of the most deteriorated patients.

Summary

Twenty-two deteriorated schizophrenics were tested to obtain psychological measurement of intellectual ability both with and without the use of sodium amytal. Deteriorated schizophrenics were defined as those: who had a known illness of at least five years duration; who were judged clinically to be deteriorated; who showed a significantly large Wechsler-Bellevue deterioration ratio; and who showed an abnormal EEG record. A test-retest group, comparable in age, length of illness, and clinical judgment was used to obtain some data on test-retest effects.

The experimental group was administered the Verbal Scale of the Wechsler-Bellevue three times—the first time conventionally, the second time under sodium amytal, and the third time under usual conditions. The test-retest group was conventionally tested twice. The primary hypothesis tested was that a statistically significant improvement in performance would be obtained with subjects under sodium amytal over non-amytal tests. This hypothesis was confirmed by the data, as these differences were statistically reliable, while the differences between the first and third testings were statistically insignificant.

A discussion of implications for psychotherapy followed indicating that the obtained results support the argument that psychotherapy might be successfully utilized with deteriorated schizophrenics, and that such patients still have a potential capacity beyond their observed behavior. The applicability of the Wechsler-Bellevue deterioration ratio to this illness was considered, the conclusion being that such use was of extremely doubtful validity. The over-all conclusion from this study was that there is a functional component in schizophrenic deterioration which may be amenable to psychological methods.

Received July 28, 1952.

References

1. Arkola, Audrey L. An experimental study of the effects of sodium amytal upon performance of the Hunt-Minnesota Test for organic brain damage. *J. clin. Psychol.*, 1947, 3, 392-396.
2. Cameron, N. The functional psychoses. In J. McV. Hunt (Ed.), *Personality and the behav-*

---

[4] Subtest analysis was also sterile both in terms of isolating differential improvement and in terms of subtest scatter. Amytal improvements were evenly distributed over all of the subtests, and the scatter remained extensive—often increasing.

ior disorders. New York: Ronald, 1944. Pp. 861-921.

3. Garfield, S. L., & Fey, W. F. A comparison of the Wechsler-Bellevue and the Shipley Hartford Scales as measures of mental impairment. *J. consult. Psychol.*, 1948, **12**, 259-264.
4. Hunt, W. L. The relative rates of decline of Wechsler-Bellevue "hold" and "don't hold" tests. *J. consult. Psychol.*, 1949, **13**, 440-443.
5. Layman, J. W. A quantitative study of certain changes occurring under sodium amytal. *Psychol. Bull.*, 1936, **33**, 745.
6. Layman, J. W. A quantitative study of certain changes in schizophrenic patients under the influence of sodium amytal. *J. gen. Psychol.*, 1940, **22**, 67-86.
7. Levy, S., & Kennard, Margaret. Unpublished study.
8. Lindemann, E. The psychopathological effect of sodium amytal. *Proc. Soc. exp. Biol. Med.*, 1931, **28**, 864-866.
9. Lindemann, E. Psychological changes in normal and abnormal individuals under the influence of sodium amytal. *Amer. J. Psychiat.*, 1932, **11**, 1083-1091.
10. Magaret, Ann, & Simpson, Mary M. A comparison of two measures of deterioration in psychotic patients. *J. consult. Psychol.*, 1948, **12**, 265-269.
11. Wechsler, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Influence of Socioeconomic Status on Wechsler Intelligence Scale for Children: An Exploratory Study

Betsy Worth Estes
University of Kentucky

The present study is a preliminary exploration toward determining whether socioeconomic level affects responses to the items of the Wechsler Intelligence Scale for Children so that test differences between upper and lower socioeconomic levels are manifest in:

1. The distribution of IQ's (means and variances).
2. The prediction value of the WISC, i.e., its correlation with an achievement test.
3. The pattern of test responses as indicated by subtests and intercorrelations between the subtests.

Subjects and Procedure

Two groups of children were studied: one of upper economic status ($N=40$), one of low economic status ($N=40$). The subjects in the upper economic group were from the university school, which is essentially a public school except for the nominal tuition charged. This school is one of several in which can be found a fairly large number of children from very comfortable homes—upper economic status. The fathers of these children are university professors, lawyers, physicians, bankers, editors, merchants, clergymen, etc. The subjects in the low economic group were from five public schools. The fathers of these children are day laborers, night watchmen, junk collectors, etc.

Of the total of 80 children, one-half were boys and one-half were girls. One-half were seven years of age, plus or minus two months, and in the second grade; the other half were ten years of age, plus or minus two months, and in the fifth grade. These two ages were selected in order to investigate whether differences in socioeconomic level effect differences in the pattern of test responses as compared with the intercorrelations of the subtests with those at seven and ten as published by Wechsler [9].

To determine socioeconomic level, the Warner-Meeker-Eells' Index of Status Characteristics was used [8]. This index uses four factors: occupation of father, source of income, house type, and characteristics of the dwelling area. The factors are rated and a weighted total of the ratings is obtained. The total is converted to social-class equivalence as upper-upper, lower-upper, upper-middle, lower-middle, upper-lower, lower-lower. Table 1 shows the social class ratings of the subjects in the study. There is no overlapping of the two groups. The subjects were selected only on the bases of sex, age, grade, and socioeconomic level.

Table 1

Distribution by Social Class Level

	Second grade		Fifth grade		
Status Index	Boys	Girls	Boys	Girls	Total
Upper Group					
Upper-Upper	0	0	0	0	0
Lower-Upper	1	0	2	2	5
Upper-Middle	6	9	5	8	28
Lower-Middle	3	1	3	0	7
Lower Group					
Upper-Lower	4	0	1	3	8
Lower-Lower	6	10	9	7	32
Total	20	20	20	20	80

The WISC was administered to all sub-

Table 2
Means, *SD*'s, Verbal and Performance Scaled Scores on the WISC

	Second grade			Fifth grade			Total group		
	M	*SD*	*t*	M	*SD*	*t*	M	*SD*	*t*
Full Scale									
Upper	114.4	9.49	4.20**	115.65	11.14	1.96	115.02	10.37	4.09**
Lower	101.8	8.99		108.15	12.45		104.98	11.31	
Verbal Scores									
Upper	72.65	8.67	3.94**	78.0	9.28	2.06			
Lower	61.65	8.98		71.0	12.06				
Performance Scores									
Upper	70.0	11.95	2.62*	68.0	11.68	1.63			
Lower	61.4	8.53		62.0	11.58				

**Significant at the 1 per cent level.
* Significant between the 1 and 5 per cent levels.

jects. The full scale of twelve subtests was used.

The second problem, that of "validity," was approached by giving all subjects the Metropolitan Achievement Test. The second-graders took Primary I Battery, Form R. The fifth-graders took the abbreviated form of the Intermediate Battery, Form R, which included tests of reading, vocabulary, arithmetic fundamentals, arithmetic problems, English, and spelling.

The third problem was investigated by means of intercorrelations of subtests of the WISC computed for the subjects in this study.

Results

I. Does Socioeconomic Level Effect Differences in Distribution of IQ's?

1. Analysis of variance using the method of repeated measurements on the same subjects, analogous to the method described by Edwards [1], shows that scores on the WISC varied with socioeconomic level, with sex, and with verbal and performance halves of the WISC.[1] Scores were analyzed in terms of sex, although, since the numbers were equal, this variable was not of primary concern in this study (Table 5, deposited with ADI).

2. Table 2 shows a difference in mean IQ for the total group, i.e., both grades combined. The mean of the upper economic group was superior to the mean of the lower economic group. The variability of the two groups as expressed by the *SD*'s did not differ significantly.

Table 3
Mean IQ and *SD* by Grade, Economic Level, Sex

	M	*SD*	*t*
Second Grade			
Upper			
Male	116.7	10.23	1.06
Female	112.1	9.04	
Lower			
Male	106.4	10.09	2.53*
Female	97.2	4.19	
Fifth Grade			
Upper			
Male	117.0	7.13	.51
Female	114.3	13.93	
Lower			
Male	111.0	11.85	.998
Female	105.3	12.38	

*Significant at better than the five per cent level.

When the total group is divided by grade (or age), it is found that in the second grade

---

[1] To reduce printing costs Tables 5, 6, 7, 8 and 9 have been deposited with the American Documentation Institute. Order Document 3837 from American Documentation Institute, c/o Library of Congress, Washington 25, D.C., remitting $1.00 for microfilm (images 1 inch high on standard 35 mm. motion picture film) or $1.00 for photocopies (6 x 8 inches) readable withhout optical aid.

(Table 2) the difference in means is again very significant, better than the 1 per cent level. There is only one subtest in which the lower group exceeds the upper group—Coding.

In the fifth grade, however, the differences in means are not significant although there is only one subtest in which the lower group exceeds the upper group—Digit Span.

Table 2 also shows the comparison of the verbal and performance scores on the WISC. The differences were significant in both verbal and performance scores in the second grade but in the fifth grade only the verbal IQ means differed.

3. Table 3 presents the differences in mean IQ for the two grades and two economic groups by sex. The difference is significant only in the lower economic group in the second grade.

Thus, in answer to the first question, differences in socioeconomic level effect differences in scores on the WISC for the total group studied. When the group is separated by grade (or age), however, these differences exist for the second grade (seven-year-olds) but are not significant for the fifth grade (ten-year-olds).

II. Does Socioeconomic Level Effect Differences in Prediction Value of the WISC?

1. The second major finding was that there is not a significant difference between the two economic groups as to predictive power of the WISC. Using Fisher's z transformation it is found that the difference between the correlations with achievement in the second grade ($r = .54$ and $r = .33$) is not significant at the 5 per cent level; nor is the difference ($r = .77$ and $r = .46$) significant for the fifth grade. Thus if achievement is used as the criterion, the WISC did not predict better for one economic group than the other.

III. Does Socioeconomic Level Effect Differences in the Pattern of Test Response as Indicated by Subtests and Intercorrelations between the Subtests?

This problem was divided into three parts:

1. In terms of intercorrelations, how does the sample used in this study compare with Wechsler's standardization data?

2. In terms of intercorrelations, how does this sample when separated by socioeconomic level compare with Wechsler's standardization data?

3. Are there differences in the intercorrelations of subtests between upper economic status and lower economic status?

1. *Comparison of whole group with standardization data.* The intercorrelations of the seven and one-half year old group were compared with those presented by Wechsler [9] and the intercorrelations for the ten and one-half year old groups with those presented by Wechsler [9]. (See Tables 6, 7, deposited with ADI.) The level of 5 per cent was computed by Fisher's z transformation as given by Edwards [1].

Those which differ are Arithmetic and Information, Similarities and Comprehension, Arithmetic and Vocabulary, Arithmetic and Digit Span, Arithmetic and Coding, Vocabulary and Digit Span, and Picture Arrangement and Object Assembly. Out of a total of sixty-six intercorrelations, nine in the seven and one-half year group differ at the 5 per cent level from Wechsler's.

At the ten and one-half year old level, the r's which differ from Wechsler's are Information and Arithmetic, Comprehension and Digit Span, Arithmetic and Vocabulary, Arithmetic and Digit Span, Arithmetic and Picture Arrangement, and Arithmetic and Coding. That is, six intercorrelations out of sixty-six differ at the 5 per cent level.

The striking observation in those tests which differ is that it is Arithmetic which so frequently enters into the intercorrelations. At seven years old, six out of the nine which differ are combinations including Arithmetic, and at ten years old, five out of the six include Arithmetic. In each of these cases, the intercorrelations in this study are lower than those presented by Wechsler. Since Arithmetic is so strongly influenced by school training, this may be a reflection of differences in educational systems.

Since only nine at the seven and one-half year level (out of sixty-six) and six at the ten and one-half year level (out of sixty-six) differ from Wechsler's data, we assume that our groups conform to the standardization data and that they are representative.

2. *Comparison of group separated on basis*

of socioeconomic level with standardization data. Each age group was separated according to socioeconomic level and the intercorrelations compared with Wechsler's standardization data for each age. Again no essential differences from the standardization group were found. (See Tables 8, 9, deposited with ADI.)

(*a*) In the upper economic group at seven and one-half years, those *r*'s which differ beyond the 5 per cent level are Information and Picture Arrangement, Information and Block Design, Information and Object Assembly, Information and Mazes, Arithmetic and Vocabulary, Arithmetic and Picture Completion, Arithmetic and Picture Arrangement, Arithmetic and Block Design, Similarities and Picture Arrangement, Vocabulary and Digit Span, Vocabulary and Object Assembly, Vocabulary and Coding, and Vocabulary and Mazes.

Those which differ in the lower group at seven and one-half years are Information and Arithmetic, Arithmetic and Vocabulary, Arithmetic and Digit Span, Arithmetic and Coding, Vocabulary and Digit Span, Digit Span and Object Assembly, Picture Arrangement and Object Assembly, and Object Assembly and Mazes.

Thus in the upper group thirteen intercorrelations out of sixty-six differ from Wechsler's; in the lower group, eight differ. Since the number of cases was so much smaller in this study than in Wechsler's group, it is assumed that there are no essential differences.

(*b*) At ten and one-half years, in the upper economic group, differences from Wechsler's are obtained in eight intercorrelations: Information and Arithmetic, Comprehension and Digit Span, Comprehension and Picture Completion, Arithmetic and Digit Span, Arithmetic and Picture Arrangement, Arithmetic and Coding, Similarities and Picture Completion, and Vocabulary and Block Design.

In the lower economic group, there are no *r*'s which differ significantly from the standardization data.

Table 4

Scaled Score Differences between Economic Groups in Subtests

Subtests	Differences in economic groups in points Second grade	 Fifth grade
Information	49	18
Comprehension	43	36
Arithmetic	12	3
Similarities	58	43
Vocabulary	53	50
Digit Span	5	−10†
Picture Completion	10	12
Picture Arrangement	51	32
Block Design	48	23
Object Assembly	37	32
Coding	−9†	16
Mazes	53	8

†No sign indicates that the upper economic group had more points. A minus sign indicates that the lower economic group was superior.

3. *Intercorrelations of subtests according to socioeconomic level.* The third comparison (Table 8, ADI) is of intercorrelations according to socioeconomic level. At seven and one-half years five *r*'s differ according to socioeconomic level: Information and Picture Arrangement, Arithmetic and Picture Arrangement, Vocabulary and Coding, Digit Span and Block Design, and Digit Span and Object Assembly.

At ten and one-half years, (Table 9, ADI) only two *r*'s differ according to socioeconomic level: Arithmetic and Digit Span, and Picture Arrangement and Object Assembly.

Table 4 shows the differences between economic groups in the subtests in standard score units. In general, it is the verbal items that show the greatest differences—Vocabulary, Similarities, Comprehension.

The results of this study, therefore, suggest that socioeconomic level effects differences (*a*) in magnitude of scores on the WISC but not (*b*) in pattern of scores and (*c*) that socioeconomic level does not effect differences in prediction value of the WISC.

Discussion

Several explanations might be suggested for the decrease in socioeconomic effects on Wechsler IQ with increase in age or grade. Does a selective factor operate so that only the brightest of the lower economic group persist to reach the same grade at the same age as the upper economic group? The duller ones may still be in the third and fourth grades and only the most capable regularly promoted.

Is it possible that the public schools are narrowing the differences occasioned by economic status in the home so that with increase in age, there is a decrease in the discrepancy between measured IQ in the economic groups? Many studies have shown a decrease in IQ with increase in age [3, 7]. However, the pupils in these studies were continuing in an extremely impoverished environment with a minimum of stimulation. The subjects in this study were in good public schools where there was constant encouragement and stimulation. This is an attractive explanation if its validity could be established.

A third explanation may be consonant with research at the University of Chicago [2, 4, 5, 6], that test items are constructed for and favor the middle and upper economic groups and penalize the lower group. Thus at an early age when the home environment is most influential, the lower group does poorly. At a later age, the middle-class influences of school, Sunday School, movies, etc., are beginning to be evidenced and may affect responses on the WISC.

Since this study suggests that socioeconomic differences as reflected in the WISC decrease with age, the WISC may become a preferred test in clinical and educational situations at older ages. Future studies might investigate this hypothesis at older ages than used in the present study.

Summary and Conclusions

Two groups of subjects, of upper and of lower economic level, 40 in each group, were selected on the bases of age, sex, grade, and socioeconomic status. One-half were seven and one-half years, plus or minus two months, in the second grade; one-half were ten and one-half years, plus or minus two months, in the fifth grade. One-half were boys; one-half were girls.

Socioeconomic level was determined by the Warner-Meeker-Eells' Index of Status Characteristics. There was no overlapping between the two groups.

The full scale of twelve subtests of the Wechsler Intelligence Scale for Children was administered. The Metropolitan Achievement Test was given to each subject.

The results were analyzed as a preliminary exploration toward determining whether socioeconomic level affects responses to the items of the WISC so that test differences between upper and lower socioeconomic levels are manifest in:

1. The distributions of IQ's.
2. The predictability of the WISC.
3. The pattern of test response as indicated by intercorrelations of the subtests (*a*) with Wechsler's standardization data and (*b*) between economic levels.

Scores on the WISC vary with socioeconomic level in respect to magnitude of score; the upper economic group was superior to the lower economic group. When the groups were divided by age or grade, the difference was significant at the seven and one-half year level, not significant at the ten and one-half year level.

The WISC did not predict better for one economic group than for the other.

Patterns of scores on the WISC did not appear to be differentiated by socioeconomic level.

Received July 21, 1952.

References

1. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
2. Eells, K., Davis, A., Havighurst, R. J., Herrick, V. E., & Tyler, R. *Intelligence and cultural differences.* Chicago: Univer. of Chicago Press, 1951.
3. Gordon, H. *Mental and scholastic tests among retarded children.* London: Board of Education, Education Pamphlet No. 44, 1923.
4. Haggard, E. A. Influence of culture background on test performance. Paper presented before Invitational Conference on Testing Problems sponsored by Educational Testing Service. New York, October 29, 1949.
5. Mitchell, J. V. Social class differences in mental test performance. Unpublished paper read at Midwest. Psychol. Ass., 1951.
6. Neff, W. S. Socio-economic status and intelligence. *Psychol. Bull.,* 1938, **35**, 727-757.
7. Sherman, M., & Key, C. B. The intelligence of isolated mountain children. *Child Develpm.,* 1932, **3**, 279-290.
8. Warner, W. L., Meeker, Marchia, & Eels, K. *Social class in America.* Chicago: Science Research Associates, 1949.
9. Wechsler, D. *Wechsler Intelligence Scale for Children.* New York: Psychological Corp., 1949.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

A Comparison of the Wechsler-Bellevue and ACE Tests on a University Counseling Center Group

Reed M. Merrill and Louise B. Heathers
University of Washington Counseling Center

As indicated by the many studies reviewed by Rabin [6], Watson [9], and Rabin and Guertin [7], the Wechsler-Bellevue Intelligence Scale (W-B) is very widely used as a measure of general ability. Relatively few studies, however, have been made regarding its validity for predicting academic performance at the college level. In this study W-B IQ's are compared with scores on the American Council on Education Psychological Examination (ACE). The ACE was selected because of its widespread use as the measure of scholastic aptitude given entering college freshmen. Two points are of interest: (*a*) the relative capacity of the two tests to predict academic achievement at the college level and (*b*) the relationship between the scores obtained on the two tests.

These relationships have been investigated in three previous studies. Anderson [1] gave the W-B and the ACE to all but two members of an entering freshman class in a woman's college. The 1940 ACE was given on entrance; the W-B, some time during the student's second semester. ACE raw scores were used in the correlations. The mean Verbal, Performance, and Full Scale W-B IQ's for this group of 112 *S*s were 116.0, 115.8, and 118.5, respectively. Sartain [8] gave the W-B and the 1942 ACE to 50 college freshmen. The sex of the *S*s was not specified. Presumably they were volunteers. ACE raw scores were used in the computations. The mean Verbal, Performance, and Full Scale W-B IQ's for this group were 115.4, 115.1, and 117.5, respectively. The measure of scholastic achievement used in both studies was the grade-point average (GPA) obtained by a student during his freshman year. Gerboth [3], in studying the relationship between Forms I and II of the W-B, gave Form I to 25 men and 25 women at some time during their college career. The median class placement of the *S*s was at the sophomore level but all classes were represented. Presumably these *S*s were volunteers. Some of these *S*s had taken some form of the ACE on entrance. The number of *S*s in the Form I group for whom ACE scores were available and the average time interval between the two testings are not known. Since various forms of the ACE had been taken by the *S*s, ACE centiles were used in the computations. The measure of scholastic achievement used was the GPA obtained during one semester, probably the semester during which the W-B was given. The mean Verbal, Performance, and Full Scale W-B IQ's for this group were 124.4, 122.1, and 125.6, respectively.

The study reported here was based on 251 male students who were part of a larger sample seen over a two-year period at a university counseling center. The total group of *S*s is described in detail elsewhere [5]. The 251 *S*s were those members of the original group who had taken some form of the ACE as entering freshmen at the University of Washington. Since the W-B was taken at the time the student came to the counseling center, the class level of the *S*s and the time interval between the two test administrations varied. When they were given the W-B, 55% of the *S*s were freshmen, 31.5% were sophomores, and 13.5% were upper-division students. Their median age at this time was 20.96 years with a *Q* of 2.26 years. The median time interval between the two tests was 11.02 months with a *Q* of 8.52 months. The measure of scholastic performance was the GPA achieved in academic subjects for the freshman year. The mean GPA[1] was 2.10 with a standard deviation of .63 and a range from 0.0 to 3.8; the mean GPA for all freshmen at the University of Washington during the period covered by the study was 2.2.[2] Table 1 presents the mean W-B IQ's and mean local entering-freshmen centiles obtained by the group. Local centiles were used in all computations since various editions of the ACE had been taken by the *S*s and these centiles were more stable from year to year for the local population than the national centiles.

Table 1

Means and Standard Deviations on the W-B and the ACE Tests of Student Client Group
($N = 248$)

	Wechsler-Bellevue IQ's		
	Verbal Scale	Performance Scale	Full Scale
Mean	117.70	118.21	120.07
σ	8.55	9.72	8.64
	Local ACE Centiles		
	Q Scale	*L* Scale	*T* Scale
Mean	52.25	47.75	49.05
σ	28.10	29.30	29.00

As may be seen in Table 2 the correlations found between GPA and both the W-B IQ's and the ACE scores tend to be quite similar in the four studies. In general the W-B Full Scale predicts college achievement as well as does the Total ACE. The Verbal Scale predicts achievement better than any of the other scales while the Performance Scale predicts less well than any of the other scales. The results on the ACE are in agreement with those reported by Berdie *et al.* [2]; ie., (*a*) the Linguistic (*L*) Scale predicts general achievement as well as the Total ACE Scale and somewhat better than the Quantitative (*Q*) Scale, and (*b*) the correlations between grades and ACE scores are consistently higher for the female than for the male groups. It is of interest that this sex difference does not appear in the W-B correlations with GPA.

Table 2

Comparison of Correlations of W-B IQ's and of ACE Scores with GPA in the Four Studies

Study	GPA, W-B IQ			GPA, ACE Score		
	V	P	FS	Q	L	T
Anderson [1]	.52	.23	.45	.36	.48	.49
Student Clients*	.58	.17	.46	.27	.39	.40
Sartain [8]	.58	.35	.53	.35	.44	.43
Gerboth [3]	—	—	.29	—	—	.41

*All *r*'s significant above the 1% level.

Table 3 gives the significance of the differences between the correlations with GPA obtained in this study. In obtaining the *t* ratios, *r*'s were transformed into Fisher's *z*'s, a statistic whose sampling distribution is normal and independent of the magnitude of the universe *r*. Since the correlations compared were all based on the same group, McNemar's [4] procedure for estimating the standard error of the difference between *z*'s was employed. This technique takes into account the intercorrelations between the measures.

Table 3

Significance of the Differences between Correlations with GPA on the Student Client Group
($N = 248$)

Comparison	$z_1 - z_2$	$\sigma_{z_1 - z_2}$	*t*
FS vs. T	+.073	.061	1.20
V vs. FS	+.166	.045	3.69*
V vs. T	+.239	.062	3.85*
V vs. L	+.251	.067	3.75*
V vs. Q	+.386	.069	5.59*
V vs. P	+.491	.071	6.92*
P vs. FS	−.325	.041	7.97*
P vs. T	−.252	.070	3.60*
P vs. L	−.240	.075	3.20*
P vs. Q	−.105	.070	1.50
L vs. Q	+.135	.067	2.01
T vs. Q	+.147	.044	3.34*

*Significant above the 1% level.

The results suggest that the time required to give the Verbal Scale is time well spent. This scale tends to predict college achievement better than does the ACE and these predictions do not appear to be affected by sex differences. In addition the scale provides the counselor with the opportunity to observe the quality of a student's behavior. Such observations are often as useful in counseling as are quantitative scores. One can speculate that academic predictions based on the Verbal Scale are best because it reflects, like GPA, both personality and ability factors.

---

[1] Since three students withdrew prior to obtaining any grades, all data involving GPA are based on 248 instead of 251 *S*'s.

[2] A grade of *C* has a point value of 2.0 at the University of Washington.

The results also emphasize the need to interpret test scores in a very conservative manner and the need to utilize data from more than one source in making predictions about general college achievement. For the highest correlation reported, .58 between the Verbal Scale and GPA, the coefficient of alienation is .81. Predictions based on a correlation of this magnitude are only 19% better than chance. Further, it should be noted that most correlations obtained between college GPA's and tests of ability have even less predictive power, particularly when predictions are made beyond the freshman level.

Table 4 presents the intercorrelations between W-B and ACE scores found in the four studies. Here there is more variability in the results, particularly with those involving the Performance Scale. The most noteworthy finding is that, in all four studies, the Verbal Scale predicts the various ACE scores as adequately as does the W-B Full Scale. In most of the comparisons the Performance Scale predicts ACE performance less well than does the Verbal Scale. Table 5 shows the significance of the differences between the intercorrelations between W-B and ACE scores found in this study.

Table 4

Comparisons of Intercorrelations of the W-B IQ's and ACE Scores in the Four Studies

Correlations	Anderson [1] (N=112)	Student Clients* (N=251)	Sartain [8] (N=50)	Gerboth [3] (N=?)
V,Q	.49	.50	.50	—
V,L	.51	.54	.72	—
V,T	.54	.61	.68	.57
V,P	.34	.43	.53	—
V,FS	—	.81	.87	—
P,Q	.31	.42	.50	—
P,L	.19	.35	.54	—
P,T	.25	.43	.55	—
P,FS	—	.80	.84	—
FS,Q	—	.55	.55	—
FS,L	—	.54	.71	—
FS,T	.48	.61	.69	.58
Q,L	.61	.49	.60	—
Q,T	—	.79	.79	—
L,T	—	.91	.94	—

*All r's significant above the 1% level.

Table 5

Significance of the Differences between Intercorrelations between W-B IQ's and Local ACE Centiles for the Student Client Group (N=251)

Comparison	$z_1 - z_2$	$\sigma_{z_1 - z_2}$	t
V,Q vs. FS,Q	−.069	.045	1.53
V,L vs. FS,L	.00	—	—
V,T vs. FS,T	.00	—	—
V,Q vs. V,L	−.055	.071	0.77
V,Q vs. V,T	−.160	.048	3.33*
V,L vs. V,T	−.105	.039	2.69*
P,L vs. V,L	−.238	.071	3.35*
P,T vs. V,T	−.249	.075	3.32*
P,Q vs. V,Q	−.101	.073	1.38
P,Q vs. P,L	+.082	.068	1.21
V,L vs. Q,L	+.068	.082	0.83
V,P vs. Q,L	−.076	.081	0.93

*Significant above the 1% level.

These intercorrelations are too low to support the assumption that, in normal persons, performance level on various kinds of ability tests should be highly similar. They are too high to support the assumption occasionally implied in test interpretation that the different scales of the two tests measure discrete abilities.

Summary

Correlations between college freshmen GPA's and the various W-B and ACE scales are reported. The findings of Anderson, Sartain, and Gerboth are reviewed and compared with the results found on 248 male student counseling center clients. Similar comparisons are made relative to the intercorrelations between the W-B and the ACE scales.

The *L* Scale of the ACE predicts the general academic achievement of college freshmen as well as the Total ACE and better than the *Q* Scale. Correlations between ACE scores and grades are higher for women than for men; this trend is not apparent on the W-B. The Verbal Scale of the W-B appears to be the best single predictor of the general academic achievement of freshmen. It also predicts the Q, L, and Total Scales of the ACE as ade-

quately as does the Full Scale of the W-B and predicts the L Scale about as adequately as does the *Q* Scale. The Performance Scale is least related to both academic achievement and the various ACE scales. It is possible that the reduced correlation between the Verbal and Performance Scales is due not only to the restricted range of talent in the college samples but also to the fact that, at this level of ability, the two scales do not measure similar factors. The data definitely support the use of the W-B as a predictor of the academic achievement of freshmen. This test, especially its Verbal Scale, predicts as well as the ACE, a test specifically designed to predict the academic achievement of college freshmen.

However, all of the obtained correlations are of such a magnitude that predictions regarding an individual student's academic achievement in freshmen college courses are subject to considerable error. Though data such as found in this study are a valuable aid in student counseling, one does well to note the variance that is unaccounted for in such results. Thus we believe that test scores are only part of the data that must be considered by a qualified counselor in helping students make decisions related to college achievement.

Received June 11, 1952

References

1. Anderson, E. E., *et al.* A comparison of the Wechsler-Bellevue, Revised Stanford-Binet, and American Council of Education tests at the college level. *J. Psychol.*, 1942, **14**, 317-326.
2. Berdie, R., Dressel, P., & Kelso, P. Relative validity of the Q and L scores of the ACE Psychological Examination. *Educ. psychol. Measmt*, 1951, **11**, 803-812.
3. Gerboth, Renate. A study of the two forms of the Wechsler-Bellevue Intelligence Scale. *J. consult. Psychol.*, 1950, **14**, 365-370.
4. McNemar, Q. *Psychological statistics.* New York: Wiley, 1949.
5. Merrill, R. M., & Heathers, Louise B. Centile scores for the Wechsler-Bellevue Intelligence Scale on a university counseling center group. *J. consult. Psychol.*, 1952, **16**, 406-409.
6. Rabin, A. I. The use of the Wechsler-Bellevue Scales with normal and abnormal persons. *Psychol. Bull.*, 1945, **42**, 410-422.
7. Rabin, A. I., & Guertin, W. H. Research with the Wechsler-Bellevue test: 1945-1950. *Psychol. Bull.*, 1951, **48**, 211-248.
8. Sartain, A. Q. A comparison of the new Revised Stanford-Binet, the Bellevue Scale, and certain group tests of intelligence. *J. soc. Psychol.*, 1946, **23**, 237-239.
9. Watson, R. I. The use of the Wechsler-Bellevue Scales: a supplement. *Psychol. Bull.*, 1946, **43**, 61-68.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

The Effect of Variations in Ambiguity on Projection in the Children's Apperception Test[1,2]

Edith A. Weisskopf-Joelson and David B. Lynn
Purdue University

How the degree of ambiguity of projective pictures affects the amount of projective material elicited by these pictures is a question which concerns the clinical practitioner as well as the theoretician. The former might be helped in the selection of effective pictures for projective techniques if he had the answer to this question. The latter might profit from gaining insight into the relationship between two important variables, namely, projection and ambiguity of the stimulus.

It is generally assumed that projection increases with increasing ambiguity. Experimental studies appear to corroborate this impression [1, 3, 4, 5, 6, 8, 10, 11, 15] with few exceptions [7, 13, 14, 16]. One of the exceptions is a study by Weisskopf [16], in which the ambiguity of TAT pictures was increased by two methods: (*a*) by incomplete tracing of the contours of the pictures, (*b*) by short exposure of the pictures. In both cases it was found that the more ambiguous pictures elicited less projection than the less ambiguous ones, when college students are used as subjects.

The present study attempts to test the validity of the above results for a different group of subjects, namely, for children.

Subjects and Materials

The subjects were 50 children, ranging in age from nine years to nine years and eleven months. All subjects attended the fourth grade of the same public school. Their IQ's on the Revised Stanford Binet Intelligence Scale, Form L, ranged from 83 to 154, with a mean of 116 and a standard deviation of 15.57.

Ten pictures were selected from Bellak's Children's Apperception Test (CAT) [2]. These pictures were reproduced by incomplete tracing of the contours, i. e., part of the contours were omitted in order to increase the ambiguity. Figure 1 shows Picture 6 of the CAT incompletely traced. The pictures used for control purposes were modifications of the same ten pictures of the CAT. They were ob-

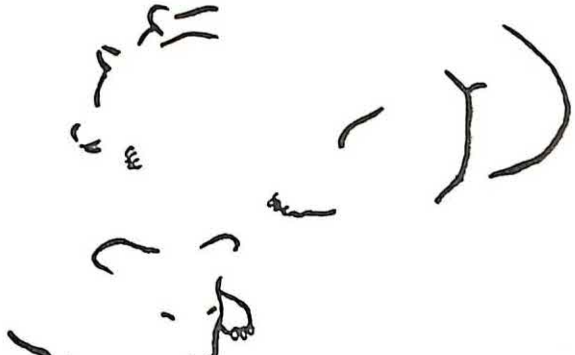


Fig. 1. Picture 6, incompletely traced.

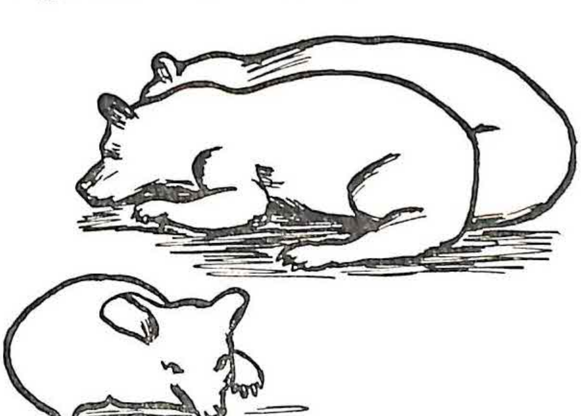


Fig. 2. Picture, 6, completely traced.

---

[1] The paper is based on a thesis presented to the Graduate School of Purdue University, Lafayette, Indiana by the junior author in partial fulfillment of the requirements for an M.S. degree in psychology. The study was directed by the senior author.

[2] This study was reported at the annual meeting of the Midwestern Psychological Association, Detroit, 1950.

tained by completely tracing the contours of the pictures, as illustrated in Figure 2. Thus, the material consisted of an experimental set of ten incompletely traced pictures and a control set of the corresponding ten completely traced pictures.

Procedure

The pictures were administered individually to each subject. The experiment was conducted in two sessions with a two-week interval between sessions. The incomplete pictures were administered in the first session and the complete ones in the second session. The order was never reversed because it can be expected that previous exposure of the complete pictures might decrease the ambiguity of the incomplete ones. However, it is reasonable to assume that the effect of repetition did not affect the results, since a previous unpublished study by Weisskopf shows that mere repetition of administration does not alter the transcendence index of incomplete pictures.

The instructions requested that the children "tell all about the picture." These instructions were used rather than the ones suggested by Bellak [2] in order to keep the procedure as similar as possible to the one followed by Weisskopf [17], who used the instructions "describe the picture." However, when using children as subjects, it was found advisable to change this latter phrase to one which has a similar meaning but is easier understood by children.

The amount of projection elicited by the pictures was measured by transcendence indices, a measure discussed in previous studies [16, 17, 18].

The transcendence index of a picture is obtained as follows: the subjects are asked to describe the picture.

> Even when the instructions call for objective description and not for fantasy, most subjects find it difficult to remain entirely descriptive. They tend to add projective material to their descriptions. *In order to obtain the transcendence index of a picture, the number of comments about this picture which go beyond pure description were counted. The transcendence index of the picture is the mean number of such comments per subject.* Pictures with low transcendence indices make it relatively easy for subjects to be factual and impersonal. Pictures with high transcendence indices make such impersonal observation difficult and, as it were, "lure" the subjects away from the objective path prescribed by the instructions and "force" him to project [17, p. 379].

The deviations from objective description were scored in terms of various categories such as the following:

> *Intraception* refers to the ascribing of emotions, desires, thoughts, fantasies, etc., to figures in the picture. . . . *Temporal Transcendence* refers to the inclusion of events which occur prior to, or after the event shown on the picture. . . . *Relationship* refers to the characterization of figures as related to other figures in the same picture. The great majority of relationship scores refer to kinship, as when one figure is described as the father, mother, son, daughter, husband, wife, etc., of another figure. . . . *Content of speech* refers to verbal statements made by pictorial figures. . . . *Atmosphere* refers to the characterization of the whole picture or parts of the picture in terms of the emotional response which it elicits in the observer. . . . *Imperative* refers to comments to the effect that a pictorial figure "should," "must," "is supposed to," act, feel, or think in a certain manner [17, p. 382-383].

The interjudge reliability, and the test-retest reliability of transcendence scoring have been previously discussed [18].

The score-rescore reliability of the junior author was found to be .98 for the complete pictures as well as for the incomplete ones.

The junior author and two additional judges scored the entire story material. Thus, each judge obtained a transcendence index for each picture. The mean of these three indices was used as the final transcendence index of the picture. The estimated reliability of this mean was found to be .96 for the complete pictures and .73 for the incomplete pictures [12]. The interjudge reliabilities of three judges ranged from .86 to .94 for the complete pictures and from .44 to .90 for the incomplete pictures.

Results

The mean transcendence index for the incomplete pictures is .60, for the complete pictures, 1.48. The difference between the two means is statistically significant well beyond the 1 per cent level. An individual comparison of each incomplete picture with the corresponding complete picture yields a difference in the same direction for eight out of ten pictures. Seven of these eight differences are significant beyond the 1 per cent level, while one is significant between the 1 and 2 per cent level.

Discussion

The results of this study are in agreement with those of the above-mentioned study by Weisskopf, which investigates the same problem, using adults as subjects. "It seems that the subjects, when confronted with an indefinite and highly equivocal stimulus, produce less fantasy material than if confronted with a more definite picture" [16, p. 412-3]. The results of both studies are in agreement with Sears's [14] findings pertaining to doll play.

The implications of the results for the development of pictorial projective techniques have been presented in the earlier study:

> Can the results of this . . . study help the originators of pictorial projective tests to select pictures of high diagnostic value? The author believes that the answer is in the affirmative. . . .
>
> Are highly ambiguous or are highly structured pictures of greater diagnostic value? By increasing the degree of ambiguity of a picture . . . [by incomplete tracing], the production of fantasy material going beyond perception, such as the ascribing of emotions to human figures, or the imaginative creation of past and future events, is decreased. . . . Such a decrease would be a loss for diagnostic purposes, because transcendent material of this kind can be used as a basis for clinical diagnosis. However, this loss is probably accompanied by a gain in another direction. Even though highly ambiguous pictures do not elicit as much material *beyond* perception, they may enable the clinician to make diagnostic use of perception itself, that is, of what and how the subject perceives. With relatively ambiguous, unstructured pictures interpretation of this latter type may be more fruitful than with relatively definite pictures. High ambiguity of pictures increases the relative influence of individual differences on perception, as the protocols of the present study show.
>
> Thus, by using such relatively structured material as the . . . pictures [of the Children's Apperception Test], the clinician loses diagnostic material of one kind and gains diagnostic material of another kind. He *loses* the possibility of basing his interpretation mainly on *what* the subject sees in the picture, since most nonpsychotic subjects show considerable agreement as to what is actually shown. But he *gains* diagnostic material consisting of *additions* to perception, such as the imaginative creation of past and future events, or the ascribing of emotions and thoughts to human figures. In the Rorschach, material of this kind is rarely produced. According to the results of this study this comparative lack of transcendence in the Rorschach may be due to the fact that the Rorschach blots are so much more ambiguous than the . . . pictures [of the Children's Apperception Test]. "Rorschach interpretation . . . is based much more on *what* and *how* the subject perceives than what he *adds* to perception" [17, p. 380]. It "is based on transcendence of sensation rather than transcendence of perception" [17, pp. 380–381].
>
> It seems that in whatever direction the ambiguity of projective stimuli is varied, diagnostic material of one kind is lost and of another is gained. . . . Thus, the present study claims a decrease with increased ambiguity for one kind of projection (transcendence) only. The study of the factors causing the negative relationship between transcendence and ambiguity may be a fruitful objective for further research [16, pp. 414-415].

Summary

The amount of projection elicited by pictures was studied as a function of the ambiguity of the pictures.

The subjects were 50 nine-year-old children. Ten pictures from the Children's Apperception Test by Bellak were traced, whereby parts of the contours were omitted to increase the ambiguity. The amount of projection elicited by these pictures was compared with the amount of projection elicited by fully traced, and thus less ambiguous pictures .The pictures with higher ambiguity elicited less projection than the less ambiguous pictures. This difference was shown to be significant well beyond the 1 per cent level.

Received June 25, 1952.

References

1. Asch, S. E., Block, H., & Hertzman, M. Studies in the principles of judgments and attitudes: I. Two basic principles of judgment. *J. Psychol.,* 1938, *5,* 219-251.
2. Bellak, L., & Bellak, Sonya S. *Children's Apperception Test.* New York: C.P.S. Co., 1949.
3. Bruner, J. S., & Goodman, C. C. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.,* 1947, **42**, 33-44.
4. Carter, L. F., & Schooler, K. Value, need, and other factors in perception. *Psychol. Rev.,* 1949, **56**, 200-207.
5. Coffin, T. E. Some conditions of suggestion and suggestibility: A study of certain attitudinal and situational factors influencing the process of suggestion. *Psychol. Monog.,* 1941, **53**, No. 4 (Whole No. 241).
6. Hertzman, M. Studies in the principles of judgments and attitudes: III. The functional

equivalence of two differently structured references. *J. soc. Psychol.,* 1940, **12,** 3-19.

7. Levine, R., Chein, I., & Murphy, G. The relation of the intensity of a need to the amount of perceptual distortion: a preliminary report. *J. Psychol.,* 1942, **13,** 283-293.
8. Luchins, A. S. Social influences in perception of complex drawings. *J. soc. Psychol.,* 1945, **21,** 257-273.
9. Luchins, A. S. The stimulus field in social psychology. *Psychol. Rev.,* 1950, **57,** 27-30.
10. McClelland, D. C., & Atkinson, J. W. The projective expression of needs: I. The effect of different intensities of the hunger drive on perception. *J. Psychol.,* 1948, **25,** 205-222.
11. McGregor, D. The major determinants of the prediction of social events. *J. abnorm. soc. Psychol.,* 1938, **33,** 179-204.
12. Mosier, C. I. On the reliability of a weighted composite. *Psychometrika,* 1943, **8,** 161-168.
13. Schonbar, Rosalea A. The interaction of observer pairs in judging visual extent and movement: the formation of social norms in "structured" situations. *Arch. Psychol., N. Y.,* 1945, No. 299.
14. Sears, R. R. Influence of methodological factors on doll play performance. *Child. Develpm.,* 1947, **18,** 190-197.
15. Sherif, M. A study of some social factors in perception. *Arch. Psychol., N. Y.,* 1935, No. 187.
16. Weisskopf, Edith A. An experimental study of the effect of brightness and ambiguity on projection in the Thematic Apperception Test. *J. Psychol.,* 1950, **29,** 407-416.
17. Weisskopf, Edith A. A transcendence index as a proposed measure in the TAT. *J. Psychol.,* 1950, **29,** 379-390.
18. Weisskopf, Edith A., & Dunlevy, G. P. Bodily similarity between subject and central figure in the TAT as an influence on projection. *J. abnorm. soc. Psychol.,* 1952, **47,** 441-445.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Sex Sequence in the Draw-a-Person Test and its Relation to the MMPI Masculinity-Femininity Scale

Samuel Granick
Washington, D. C.

and Leon J. Smith
St. Petersburg, Florida

Subjects are expected to draw a figure of their own sex as their first response to the examiner's direction to "draw a person" [2, 5, 6]. Levy [5] reports that of 5000 adults examined, 87 per cent drew their own sex. A study of 80 college students, ranging in age from 18 to 58 years with a median of 26, showed similar results, 76 per cent of the subjects drawing their own sex as their initial response [3]. With reference to children between the ages of 5 and 12 years, Jolles [4] found that both the boys and the girls drew their own sex predominantly when taking the H-T-P Test. Data obtained from a group of Serbian refugee children [1], on the other hand, indicated that while boys showed a preference for drawing males, girls under 10 tended to draw their opposite sex predominantly. The girls of 11 and 12 years of age, however, showed a preference for their own sex. How adolescents tend to respond has not as yet, to the authors' knowledge, been reported in the literature.

What the drawing of the same or opposite sex as the first response to the Draw-a-Person test implies clinically for an individual and generally for various age and sex groups appears never to have been systematically explored. Machover [6] and Buck [2] emphasize that in drawing the human figure an individual is probably making some kind of portrait of himself. In Machover's words, "The process of drawing the human figure is for the subject, whether he realizes it or not, a problem not only in graphic skill, but one of projecting himself in all of the body meanings and attitudes that have come to be represented in his body image." She goes on to mention that empirically she has found "evidence of some degree of sexual inversion . . . in records of all individuals who drew the opposite sex first" [6, pp. 101, 102].

The study herein reported is not intended as a test of Machover's entire hypothesis. Indeed that would be exceedingly difficult to accomplish since defining "some degree of sexual inversion" might prove elusive. Intended, however, is an investigation of the incidence of self-sex drawings as the initial response to the Draw-A-Person test on the part of a group of adolescents and a comparison of this with results obtained on a standard masculinity-femininity scale.

Method

The subjects for the experiment consisted of 185 male and 386 female students in the general psychology course during the fall semester of 1950-51 at the Florida State University. Since almost all students at the university tend to enroll in this course during their undergraduate years, this sample is probably representative, in terms of intelligence and cultural background, of the more than 6000 students who were enrolled at the school during that period and who, in turn, were mainly from the various sections of the state of Florida. The ages of the subjects were not recorded, but since almost all the students were in their sophomore year, it is probably safe to assume that few were above 22 or below 17 years of age. The vast majority ranged in age between 18 and 20 years.

A booklet of four pages was prepared, two of the pages to be used for the drawings and the other two containing the 60 questions or statements of the Minnesota Multiphasic Personality Inventory Masculinity-Femininity Scale to which answers of "true" or "false" were to be given. The tests were administered by the junior author to the subjects in 16 groups of approximately 40 students each. When each subject had a booklet, the group was instructed as follows. "On the front page you are to draw a person. I am not interested in your ability to draw, but rather in how you attempt to draw a person. Please do not turn the pages of the booklet until told to do so." When all subjects had completed the first drawing, they were further instructed as follows. "Now turn to the next page. On this sheet you are to draw a person of the opposite sex to the one you have just drawn. If you drew a male on the front page, draw a female here. If you drew a female, now draw a male." When this drawing had been accomplished, the following directions were given in the administration of the M-F Scale of the MMPI. "Please place the letter *M* for 'male' or the letter *F* for 'female' in accordance with your sex in the space provided at the top left-hand corner of the third page. Now circle *T* for 'true' and the letter *F* for 'false' to these statements as they apply to you." Invariably students questioned whether or not they should append their names to the booklets. They were told it was not necessary to do so unless they wished to. Mention was also made that their responses would in no way affect their grades in the course.

Results

Of the 185 males, 163 or 88.1 per cent drew their own sex first. Similarly, the majority of the female subjects drew their own sex first (252 or 65.3 per cent of 386), but the proportion was significantly lower ($CR = 6.7$, giving a p which is less than .01). The *SE*'s of the percentages were relatively low in each case (2.40 for the males and 2.42 for the females), thus indicating that the occurrence of the self-sex drawings for both groups was beyond chance expectation. These results agree fairly closely with Holtzman's report [3] showing that 85 per cent of the males ($N = 40$) and 67.5 per cent of the females ($N = 40$) drew their own sex as their first responses to the Draw-a-Person test. Generally, therefore, the sex sequence of the human-figure drawings differentiated the sexes on a group basis.

Table 1

MMPI M-F Scale Scores for Subjects

Subjects	*N*	Mean	*SD*	*t*	*p*
Males	185	26.30	4.02	29.6	<.001
Females	386	37.26	4.50		

The results of the MMPI M-F Scale were found also to separate the two sex groups as shown in Table 1.

It should be noted, furthermore, that relatively little overlapping occurred between the distributions of the two sets of scores.

Table 2

Sex Sequence in the Draw-a-Person Test and MMPI M-F Scale Scores

Subjects	No. Own Sex First	M-F Mean	M-F *SD*	No. Opposite Sex First	M-F Mean	M-F *SD*	*t*
Males	163	26.30	4.22	22	26.22	5.10	.07
Females	252	37.64	4.16	134	36.67	4.90	.18

The question now arises as to what relationship, if any, may exist between the incidence of self-sex drawings as a first response to the Draw-a-Person test and the ratings achieved on the M-F Scale. Calculation of the point biserial correlations for both sexes indicated that this relationship is probably of a zero order ($r_{\text{pbi}} = .009$ for males and $r_{\text{pbi}} = .105$ for females). A comparison of the M-F Scale means of the subjects of each sex who drew the same sex first with the means of those who drew the opposite sex first, as shown in Table 2, further substantiated this. In addition, the differences between the mean M-F Scale scores of those drawing either of the sexes first were in the direction of the sex of the subjects rather than of the sex drawn first. Thus a comparison of the scores of the males and females drawing the male figure first yielded a t of 19.9, showing a highly significant difference between the M-F Scale means. Similarly a t of 10.2 was obtained from the comparison of the M-F Scale scores of those subjects in both sex groups who drew the female figure first. Finally, a t of 8.9 was found between the M-F Scale scores of both sexes for those drawing the opposite sex figure first.

Discussion

Evident from the data is the fact that both the Draw-a-Person test and the MMPI M-F Scale serve to differentiate the male and female groups quite definitively. Apparently the tendency to draw the same sex first, however, is far stronger in men than in women. In these findings the present study is in essential agreement with results reported by other investigators. Added only is the fact that what is found among young children [4] and adults [3, 5] is also clearly the case with older adolescents attending college.

Not so clear, on the other hand, is the significance of the self- or opposite-sex drawings as the initial response to the Draw-a-Person test. Insofar as the MMPI M-F Scale is a measure of psychosexual identification, the data would tend to refute the hypothesis of Machover [6] and Buck [2] suggesting that sexual inversion may be involved in the choice of the opposite sex as the first response to the request of the examiner to "draw a person."

Normal cultural factors are probably among the important determinants of the trend noted in this and other reports for people to draw their own sex predominantly on the Draw-a-Person test. Those subjects who fail to follow the trend may, of course, be responding to different cultural influences, or to a variety of internal stimuli, which lead to the association of opposite sex as their initial reaction. Empirically Machover [6] and Buck [2] have found reason to believe that one's drawing of the human figure is essentially a self-portrait. But this may not be as generally true as they lead one to suppose. It is logical to suppose that the sex of the person one is inclined to draw as one's first response to the Draw-a-Person test may be related to internal and external stimuli far removed from psychosexual identification and self-portraiture. This is not to suggest, on the other hand, that in many instances conscious or unconscious portrayal of one's body image is not involved. Evident, however, is the need to investigate the fantasies and environmental factors affecting an individual at the time he is responding to the Draw-a-Person test.

Summary and Conclusions

The Draw-a-Person test and the MMPI M-F Scale were administered to a group of 185 male and 386 female college students taking the general psychology course. The sex of the figure drawn as the first response was noted in each instance and compared, in turn, with the scores made on the masculinity-femininity test. An analysis of the data yielded the following conclusions.

1. Male and female college undergraduates tended to draw their own sex predominantly when directed to "draw a person." The incidence of same-sex drawings was significantly greater among males than among females.

2. The MMPI M-F Scale significantly differentiated between the male and female subjects.

3. No relationship was found between the sex sequence of human figures drawn and scores on the masculinity-femininity measure.

Received June 4, 1952.

References

1. Baumgarten, F., & Tramer, M. Kinderzeichnungen in vergleichend psychologischer Beleuchtung. (Child drawings in the light of comparative psychological study.) *Z. Kinderpsychiat.*, 1943, **9**, 161-220. (Abstracted by: Florence L. Goodenough, & D. B. Harris. Studies in the psychology of children's drawings: II. 1928-1949. *Psychol. Bull.*, 1950, **47**, 369-433.)
2. Buck, J. N. The H-T-P Test. *J. clin. Psychol.*, 1948, **4**, 151-159.
3. Holtzman, W. H. The examiner as a variable in the Draw-a-Person test. *J. consult. Psychol.*, 1952, **16**, 145-148.
4. Jolles, I. A study of the validity of some hypotheses for the qualitative interpretation of the H-T-P for children of elementary school age: I. Sexual identification. *J. clin. Psychol.*, 1952, **8**, 113-118.
5. Levy, S. Figure drawing as a projective test. In L. E. Abt & L. Bellak, *Projective psychology.* New York: Knopf, 1950. Pp. 257-297.
6. Machover, Karen. *Personality projection in the drawing of the human figure.* Springfield, Ill.: Charles C Thomas, 1949.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

A Note on the "Tree and Trauma" Interpretation in the HTP[1]

Murray Levine and Eugene H. Galanter[2]
University of Pennsylvania and VA Mental Hygiene Clinic, Philadelphia

In his manual Buck [1, p. 390] suggests that blemishes drawn on the tree, such as scars, knotholes, broken branches, points of unusual shading or reinforcement, or weak or dead branches, may be representative of traumata which have occurred in the life of the person making the drawing. Furthermore he suggests [1, p. 340] that the ratio between the height of the traumatic indicator and the total height of the tree will yield the approximate age, plus or minus one year, of the occurrence of the trauma. The suggested interpretation has practical importance in terms of diagnosis, and it also contains implications for the theory of projection in expressive techniques. Specifically it would mean that the tree drawing has some relationship to the self concept as the self concept is projected into a "life space." It is then of considerable interest to treat the interpretative suggestion as an hypothesis and to attempt an objective verification of that hypothesis.

The clinical evidence related to this suggestion is seemingly derived from the *post hoc* examination of the individual's life history for incidents considered traumatic by the individual. The interrogation of the subject involves suggesting the age of traumatic occurrence to him and his response is therefore *not* independent of the indicator in the drawing. Furthermore the criteria for a traumatic occurrence are not specified and any unusual occurrence mentioned by the subject is accepted as a valid "trauma." An objective test of the validity of the interpretation requires that both the trauma and its date of occurrence be specified independently of the drawing.

The writers obtained drawings of a tree from 27 hospitalized, paraplegic veterans assuming that the severity of the injury and the circumstances under which it occurred, as well as the complete reorientation to life the injury requires, make it eligible as a trauma. In addition one can specify the exact date of occurrence of the trauma. Both the trauma and the time of occurrence are in this way specified independently of the drawing. The instructions were simply "Draw a tree." The population ranged in age from 18 to 40, the median age being 26 years and the duration of injury ranging from 6 months to 10 years, with a median duration of 15 months.

The presence of traumatic indicators on the tree was independently rated by the authors according to the general criteria stated by Buck. The raters agreed as to presence or absence of indicators in 82 per cent of the cases. The degree of agreement is significantly greater than chance agreement of 50 per cent. Of the 27 drawings rated, only 7 contained unequivocal traumatic indicators. The fact that only a small percentage of this obviously traumatized population drew indicators which could be reliably rated as such suggests that either the hypothesis has dubious validity, or that other unknown personality factors contribute to the appearance or nonappearance of such indicators on tree drawings. The age of occurrence of the trauma was calculated in each of these 7 drawings and compared with the actual age of traumata as obtained from the patient's records. The correspondence was poor. None of the cases fell within the two-year limit suggested by Buck. However, a rank-order correlation between the estimated age of occurrence obtained from the drawing and the actual age of occurrence obtained from hospital records proved to be plus .73; significant for the 7 cases at approximately the .05 level of confidence. In addition to the few cases, confidence in the size of this coefficient is lessened by the presence of split ranks. These results, however, suggest that the hypothesis may have some degree of validity in terms of a broad approximation of the time of injury, but statements concerning more exact localization in time are not supported by our data.

---

[1] Reviewed in the VA and published with the approval of the Chief Medical Director. The statements and conclusions published by the authors are the result of their own study and do not necessarily reflect the opinion or policy of the VA.

[2] The authors wish to thank Dr. L. J. Hatterer of the Kingsbridge VA Hospital, New York, who was instrumental in helping to gather the reported data.

Since only a small percentage of this traumatized population spontaneously drew traumatic indicators, future research will require dealing with a very much larger parent population in order to obtain sufficient cases to test the hypothesis adequately. Verification of the interpretation, even in the form suggested by the present results, would seem to have interesting theoretical consequences for the general concept of projection in expressive techniques, and specifically for the notion that the self concept can be studied by this and similar methods. Others who may be working with similarly traumatized individuals might profitably repeat the study with larger samples.

Received June 6, 1952.

Reference

1. Buck, J. N. The H-T-P technique: A qualitative and quantitative scoring manual. *J. clin. Psychol.,* 1948,4, 317-396.

NEW BOOKS AND TESTS

Books

Ashby, W. Ross. *Design for a brain.* New York: John Wiley & Sons, 1952. Pp. ix + 260. $6.00.

This is another excellent addition to the list of distinguished treatises on cerebral mechanisms that Wiley has been publishing. The problem that Ashby has set himself is "to deduce from the observed facts" what are the necessary properties of a nervous system that acts "at once mechanistically and adaptively." The book is well written, the thesis clearly and tightly developed, and there is much more use of and integration with specific biological findings than one is apt to get in theoretical works of this sort. It remains, however, of necessity a difficult book which merits a larger audience than it will probably get. Its appeal will be greater to students of the logic of science and of system construction than it will be to practicing clinicians. The appendix contains a 50-page restatement of the thesis in mathematical form.—*W. A. H.*

Baumgarten, Franziska (Ed.) *La psychotechnique dans le monde moderne.* C. R. du IX^{e} Congres International de Psychotechnique. Paris: Presses Universitaires de France, 1952. Pp. xi + 630. 2,000 fr.

This is the report of the 9th International Congress of Psychotechnology held in Berne, Switzerland, in 1949, fifteen years after the last previous Congress. In addition to the welcoming addresses, and a report by the General Secretary of the preceding fifteen years of the work of the International Association of Psychotechnology, the volume includes over 100 papers presented by persons representing 23 different countries. More than half of the papers are in French, the others are in German, English, or Italian. These papers are presented under fourteen topical headings representing a wide array of problems and applications. Although containing but few detailed research reports, the volume offers an unusual opportunity to become acquainted with the current thinking and practice of our colleagues in other lands. Only seven of the papers were by Americans!—*E. L. K.*

U. S. Federal Security Agency. Public Health Service. National Institute of Mental Health. *Mental health and motion pictures.* Publ. Hlth Serv. Publ. No. 218. Washington: U. S. Government Printing Office, 1952. Pp. 124. $.30.

For anyone who has the responsibility of publicizing or training in the various aspects of mental health this guide to 106 films is a positive must. The films are grouped under Child and Family, Marriage, Mental Health and Schools(for teachers, and for students), Mental Health Problems, Physiology of Human Reproduction, and Miscellaneous. The synopses clearly present the description of the film, its aims, and other information pertinent to its use. The format is excellent, giving running time, distributor, television information, followed by ample space for personal comments by the local viewer. The "Suggested Audiences" contains valuable information. Care has been taken not only to state the audience for which the film was planned but also, in the case of films whose presentation needs more careful controls, the type of control needed such as the presence of a trained professional worker. The book contains a list of agencies designated as State Mental Health Authorities, names and addresses of distributors, and an excellent index.——*B. M. L.*

Frosch, John. (Ed.) *The annual survey of psychoanalysis.* Vol. I, 1950. New York: International Universities Press, 1952. Pp. xvi + 556. $10.00.

"This is the psychoanalytic literature of 1950, all of it. . . . " With this opening statement, a new survey venture in psychoanalytic literature was launched. Apparently other annual surveys are to follow. The publication aims at abstracting all the periodical literature of the field and many of the books for a given year; it does not aim at critical reviews. Eleven of the twelve chapters deal systematically with articles in the field, while the last chapter abstracts 12 of the 42 books published during the year 1950. Some idea of the subject organization can be gained from several chapter titles: History, Methodology, Clinical Studies, Child Analysis, Psychoanalytic Therapy, Applied Psychoanalysis, etc. Many contributors are listed as abstractors, and most of the chapters have different authors. The job done in this survey is tremendous, but even more important is the quality of most of the abstracts. Comparing the abstracts with some of the complete articles creates great respect for the

---

Note.—The reviews were prepared by the Editor and the Associate Editors, who may be identified by their initials.

compilers of this publication. Not infrequently is an article improved in readability and clarity by the abbreviation; writers in the psychoanalytic field are not as a rule adherents of the Law of Parsimony. The *Annual Survey of Psychoanalysis* is a welcome addition to psychological and psychiatric literature. Psychoanalysis will probably win more friends and more understanding through this volume than through the original material it attempts to abstract.—*M. K.*

Hirsh, Ira J. *The measurement of hearing.* New York: McGraw-Hill, 1952. Pp. ix + 364. $6.00.

The fields of experimental psychology and clinical audiometry are nicely integrated in this comprehensive and useful book on hearing. The introductory chapters give enough background on psychophysics, the physics of sound, and electroacoustic devices to serve the needs of readers who have at least a minimum acquaintance with the physiology and psychology of audition. The major problems taken up in the remaining chapters include the measurement of the hearing of pure tones and of speech, masking, fatigue, differential sensitivity, binaural hearing, bone conduction, and the use of conditioned response methods in audiometry. Because hearing loss is often related to a person's broader social adjustments, clinical psychologists have good reason to maintain more than a casual acquaintance with audiometry. They will find Hirsh's book an excellent resource.—*L. F. S.*

Kornitzer, Margaret. *Child adoption in the modern world.* New York: Philosophical Library, 1952. Pp. xiii + 403. $4.50.

The greater part of this book, written by the press officer of a group of British adoption societies, concerns present practices with detailed description of legal procedures in Great Britain. There is also some routine discussion of adoption problems generally, selection of adoptive parents, telling the child, the unadoptable child, and so on. Adoption first became legal in Great Britain in 1926; in 1949, seventeen thousand adoption orders were made. About one-third of the book discusses adoptive practices in other parts of the British Commonwealth, and brief reference is made to the situation in other countries. Two chapters are devoted to the United States. These summarize the legal situation in the various states and give some over-all comparison with that in Great Britain, but contain chiefly a "mosaic of brief quotations . . . because only in this way is it possible for a British writer to present, primarily for British readers and without causing offence to American ones anything like a valid picture of what goes on in the American adoption field today."—*A. R.*

Kuhlen, Raymond G., & Thompson, George G. (Eds.) *Psychological studies of human development.* New York: Appleton-Century-Crofts, 1952. Pp. xiii + 532. $3.50.

Currently, source books for introductory courses in psychology are finding their way into the classroom in increasing numbers. This paper-bound volume is a collection of 71 abridged and edited primary source materials in the general area of developmental psychology—here defined as extending from infancy through senescence. The subject matter covered includes physical factors in development, learning, intellectual development, social-cultural conditions, interest patterns, vocational orientation, and personal-emotional adjustment. The papers, originally published between 1928 and 1951, represent not only different research areas, but different methodologies, with the result that the studies selected will be considered of uneven quality by many readers. Although many studies now considered "classics" are included (e.g., Gesell, Jones & Conrad, Terman, Hartshorne & May, and Levy), it is inevitable that most readers will find some of their particular favorites missing. Although in general the abridgment and editing are done with care, it is questionable whether some of the more extensive abbreviations are wise (as, the reduction of Levy's book, *Maternal Overprotection,* to eight pages). These problems intrinsic to source books, however, should not detract importantly from the usefulness of this volume as a teaching aid in introductory courses in developmental psychology.—*A. M.*

Wolfle, Dael. (Chairman) *Improving undergraduate instruction in psychology.* New York: Macmillan Company, 1952. Pp. vii + 60. $1.25.

This is a 60-page report of a study group which attacked the problems of the objectives and the curriculum of undergraduate instruction in psychology. The members of the committee are well known and respected by their colleagues and the report will carry weight. A recommended curriculum with core, special interest, advanced, and integrative courses is presented for the typical liberal arts undergraduate. It is frankly a psychology- or content-centered curriculum rather than a student- or problem-oriented one. The educational objectives of personal growth and social adjustment are considered but rejected in favor of knowledge, habits of thought, and attitudes and values. The report states further that "it is probably desirable to avoid direct frontal efforts to modify student's habits and attitudes." Little space is given to indirect and student-centered means of dealing with attitudes and values that do not employ the methods the committee rightly banned such as "pressure," disregard for student rights, or reformation. Although a traditional and conservative educational climate pervades the report, there is an occasional hint that in committee discussion the realities of student motivation, individual differences, and present societal needs occasionally arose. The published report, however, does not seem to give much significance to these influential factors which will determine how well their educational objectives can be met. This report can render a real service if it provokes debate and research on both the potential role of psychology in a liberal education and the means by which this can be realistically accomplished with our present student body. It will serve also as a target of attack for those who feel along with this reviewer that the psychologist has a special responsibility and competency adroitly to assist students to meet their personal and social needs, and who also maintain that this is a primary educational goal and basic to a liberal education.—*F. McK.*

Tests

Bennett, George K., Seashore, Harold G., & Wesman, Alexander G. *Differential Aptitude Tests.* Manual (2nd Ed.), pp. 77 ($2.00). New York: Psychological Corp., 1952.

The *Differential Aptitude Tests* have been provided with an entirely new manual which replaces the original loose-leaf edition of 1947 and its five supplements issued between 1948 and 1951. In addition to the inevitable directions for administration, scoring, and use, the new manual contains unusually comprehensive data about the tests. Norms are now based on groups of 2100 to 7400 for Form A, and of 350 to 1075 for Form B, for each sex and for each grade from 8 to 12, separately. The validity of the *DAT* is described in terms of its prediction of subsequent course grades, achievement test scores, and postschool careers. There are about 20 pages of tables of validity coefficients. Also presented with more than minimum adequacy are reliabilities, standard errors of measurement, and the intercorrelations of the tests. Even more detailed research data have been deposited with the American Documentation Institute so as to be available to technically trained users. The *Manual* can serve as model for other test constructors, and as an adjunctive text for courses in measurement.—*L. F. S.*

Books Received

Adler, Leta M., Coddington, James W., & Stewart, Donald D. *Mental illness in Washington County, Arkansas: Incidence, recovery, and posthospital adjustment.* Little Rock, Ark.: Univer. of Arkansas, 1952. Pp. 74.

Brennan, Robert E. *General psychology.* (Rev. Ed.) New York: Macmillan, 1952. Pp. xxxii + 524. $5.50.

Coombs, Clyde H. *A theory of psychological scaling.* Engng Res. Bull. No. 34. Ann Arbor, Mich.: Univer. of Michigan, 1952. Pp. vi + 94. $1.75.

Eisler, Robert. *Man into wolf, an anthropological interpretation of sadism, masochism, and lycanthropy.* New York: Philosophical Library, 1952. Pp. 286. $6.00.

Lindner, Robert. *Prescription for rebellion.* New York: Rinehart, 1952. Pp. 305. $2.75.

Maslow, Paul. *Powers of the mind.* Brooklyn, N. Y. (16 Court St.): Author, 1952. Pp. 288. $3.50.

Meerloo, Joost A. M. *Conversation and communication, a psychological inquiry into language and human relations.* New York: International Universities Press, 1952. Pp. x + 245. $4.00.

Palmer, Harold. *The philosophy of psychiatry.* New York: Philosophical Library, 1952. Pp. ix + 70. $2.75.

Powys, John Cowper. *The inmates.* New York: Philosophical Library, 1952. Pp. 318. $4.50.

Helping parents understand the exceptional child. Proc. Annu. Conf. Educ. Except. Child. Langhorne, Pa.: The Woods Schools, 1952. Pp. 42.

Journal of Consulting Psychology
Vol. 17, No. 2, 1953

Rigidity and Shock Therapy of Psychotics: An Experimental Study

Maxwell S. Pullen
Ohio University

and Ross Stagner
University of Illinois

Convulsive shock therapy has, in a little more
ecade, become a major tool in the psy-
reatment of mental disorders. Particu-
those psychotics who are inaccessible
otherapy, it has become practically a
prescription in many hospitals. Yet it
o exaggeration to say that today we
latively little about its actual behav-
cts (outside of amnesia); and certain-
ve no generally accepted theory as to
rlying processes involved. Psycholo-
psychiatrists urgently need a single formulation which will clearly relate the symptoms and dynamic structure of the individual patient to the effects of shock therapy, if this is to be used with maximum efficiency.

One such formulation is that shock therapy modifies rigidity among psychotics [see, e.g., Billig and Sullivan, 1]. This implies (*a*) that rigidity is a major component of the psychotic syndrome, and (*b*) that one effect of shock therapy is to change this component. To verify this hypothesis or disprove it, we shall need agreement as to the meaning of rigidity, and a good measure of its extent before and after treatment.

The notion that psychotics, especially schizophrenics, are unusually "rigid" has considerable appeal. The schizophrenic is likely to show stereotyped behavior, inability to change when the situation calls for change. His beliefs are set and resistant to logic. It is difficult to get him to "see things differently." However, when we look at the experimental literature on the measurement of rigidity, we encounter complications. There are marked differences of opinion as to whether rigidity is a unitary phenomenon, and even whether "rigid" behavior as defined by the experimenter fits the psychotic syndrome just described.

Since the literature has recently been analyzed by Cattell [2], Cattell and Tiner [3], Fisher [4], Luchins [6], Rokeach [8], and others, we shall not attempt another survey. Some of the conclusions offered by Fisher will be accepted as a starting point for a description of our specific research design:

1. In an intellectually homogeneous group, individual differences in rigidity exist.
2. Individuals with organic pathology are more rigid than individuals without such pathology.
3. Certain kinds of schizophrenics seem to be more rigid than other kinds.
4. Isolated persons seem to be more rigid than non-isolated persons who are otherwise comparable.
5. When different kinds of rigidity measures are applied to the same group, there frequently are real differences in the results given by the various measures.

On the basis of our study of the literature on rigidity and on shock therapy, we concluded that there was ample justification for a study designed to test the hypothesis that convulsive shock therapy is followed by a reduction in the rigidity of the patient's behavior.

Method

The factor-analytic studies of Cattell and others indicate that there may be more than one kind of rigidity. Certainly there are differing definitions of the phenomenon. At the outset, therefore, we elected to define rigidity as "a generalized tendency to cling to or perseverate with a previously-made response, in lieu of switching to a response which is more appropriate to the prevailing experimental conditions." The choice of tests was then made in conformity with this definition, guided also by the general intelligence level and cooperativeness of our proposed subjects.

We have adopted four tests from the battery used by Cattell and his students. One of these is a motor test (creative effort), while three are perceptual (hidden objects; riddles; and flicker fusion threshold). One test was borrowed from V. W. Grant (hidden words). Four new tests were devised for the study: changing colors; changing figures; changing pictures; and flicker fusion overlap. Also used in the battery was the Short Form Wechsler-Bellevue scale. The following brief descriptions identify each test:

1. Changing colors: one set of 20 cards varied by minute gradations from a saturated red to a saturated orange, and another set of 26 cards varied from blue to green. The subject saw the cards within a black exposure box, uniformly illuminated, and presented in series.[1] He was told that the color would change, but that he was simply to name each color as he saw it. Each *S* did four runs, one in each direction with each set of cards. The measure was the amount of *overlap* (persisting with a response past the point at which he had abandoned it when going in the other direction). Thus, he may have switched from "red" to "orange" on card 15, but from "orange" to "red" at card twelve. The overlap of three cards presumably signifies rigidity in failing to change response.

2. Changing figures: 20 cards ranged from a perfect equilateral triangle to a perfect circle; and 22 cards ranged from a perfect square to a four-pointed star. Measure as in test 1.

3. Changing pictures: 11 cards changing gradually from the face of a pirate to a silhouette of a rabbit (adapted from Leeper, [5]); 13 cards changing from young woman to old woman (adapted from Leeper); 11 cards changing from an old man fishing to a young boy (adapted from Luchins). The score on this series is simply the number of the card at which the new percept is noted.

4. Creative effort: subject writes "ready" normally, for one minute; then writes the same word in reverse script, for one minute. Next he writes "237" normally and in reverse; and then his own first and last name, normally and in reverse. Score is given by $(x - y/x)$, in which x is the number of letters normally and y the number in reverse stroke. (Adapted from Cattell.)

5. Hidden objects: five mimeographed puzzle pictures. Score is number of hidden objects located in one minute per picture (Cattell).

6. Hidden words: Subject is shown a card with the letters: $\underline{\text{W}}$ I L L $\underline{\text{O}}$ W $\underline{\text{L}}$ E A $\underline{\text{F}}$ as marked. He is shown that the underscored letters give the name of an animal. After two more marked samples, he is given the test of 12 unmarked cards, at the rate of 30 seconds per card. The score is the number of animal names located within the time limit.[2]

7. Riddles: after a few samples to be sure that *S* understands, he is given 12 riddles, each to be solved within 30 seconds. Score: number solved.

8. Flicker fusion threshold: *S* was partially dark adapted with dark goggles during the riddles test. He is now asked to look into a black shadow box at a square of milk-glass, behind which is a Strobotac, set at 3700 RPM. *E* then slows the Strobotac down until *S* observes flicker, and goes all the way down to 600 RPM, then back. Score: mean reading of 8 ascending and 8 descending determinations of threshold (not including one practice run each way). Low fusion threshold is assumed to indicate rigidity.

9. Flicker fusion overlap: the mean of the reported thresholds for the 8 "descending" runs is subtracted from the mean for the 8 "ascending" runs, on the hypothesis that clinging to the percept once established indicates rigidity.

10. The weighted scores on Comprehension, Arithmetic, and Similarities of the Wechsler-Bellevue were added to give an intelligence measure for inclusion in the battery.

Subjects

The Experimental Group consisted of 35 psychotic patients from the Danville VA Hospital who were designated to receive convulsive shock therapy (electric or insulin). The Control Group included 25 psychotic patients from the Danville VA Hospital; it was chosen to be similar to the Experimentals in the following respects: (*a*) average age; (*b*) diagnostic categories represented; (*c*) duration of illness; (*d*) degree of incapacitation; (*e*) amount of preshock rigidity, as measured by our battery.

Each experimental subject was tested not earlier than 72 hours prior to his first shock treatment, and again, not later than 72 hours after his last shock treatment. Control subjects were tested as they were chosen, and later

---

[1] All testing was done individually by Mr. Pullen. We wish to thank the administration (Dr. O. D. Timm, Chief of Professional Services, Dr. L. A. Pennington, Chief Psychologist, and Dr. L. E. [illegible], Manager) at the Danville VA Hospital for their cooperation on this study.

[2] We wish to thank Dr. V. W. Grant, Hawthornden State Hospital, Macedonia, Ohio, for permission to use this test.

Table 1

Intertest Correlations

Test	2	4	5	6	7	8	9	10
1. Changing colors	.59	.31	–.41	–.15	–.03	–.01	.22	–.01
2. Changing figures		.25	–.38	–.23	–.15	–.07	.14	–.17
4. Creative effort			–.32	–.32	–.09	–.06	.08	.04
5. Hidden objects				.65	.57	.66	–.52	.22
6. Hidden words					.57	.28	–.24	.50
7. Riddles						.18	–.41	.54
8. Flicker fusion threshold							–.21	.00
9. Flicker fusion overlap								–.27
10. Wechsler-Bellevue								

retested at such intervals as to make the test-retest interval comparable for the two groups.

Statistical Analysis

Our first purpose was to establish a composite single score for preshock rigidity for each patient. To achieve this we proceeded as follows: the nine tests in the battery were intercorrelated (Pearson *r*) to give the matrix shown in Table 1. (The "changing pictures" test, Test 3, was dropped from the statistical analysis because on retest virtually every control subject recognized it immediately and made the switch purely on a memory basis.) All 60 subjects (experimental and control) were used for these correlations.

Table 2

Reference Factor Loadings

Test	Factor				
	I	II	I^2	II^2	h^2
1	–.49	.61	.24	.37	.61
2	–.53	.51	.28	.26	.54
4	–.34	.20	.12	.04	.16
5	.80	.19	.64	.04	.68
6	.77	.41	.59	.17	.76
7	.69	.59	.48	.35	.82
8	.19	.09	.04	.01	.12
9	–.51	–.24	.26	.06	.32
10	.43	.30	.18	.09	.27

A Thurstone centroid factor analysis was made of Table 1. Two factors emerged, which upon rotation were identified as a "rigidity" factor and an "intelligence" factor (Tables 2 and 3 and Fig. 1). Heavily loaded for rigidity were Tests 1 and 2, with 4 having a moderate positive loading, 5 moderate negative.

Table 3

Rotated Factor Loadings

Test	Factor*				
	I′	II′	I'^2	II'^2	h^2
1	.78	–.12	.61	.01	.62
2	.73	–.19	.53	.04	.57
4	.38	–.20	.14	.04	.18
5	–.40	.79	.16	.62	.78
6	–.22	.87	.05	.76	.81
7	–.04	.89	.00	.79	.79
8	–.06	.20	.00	.04	.04
9	.18	–.56	.03	.31	.35
10	–.08	.52	.01	.27	.28

* I′: "Rigidity."
II′: "Intelligence."

We next proceeded to compute *z* scores for individual subjects on the rigidity factor. This was done as follows: Thurstone has pointed out that $FP = S$, where S is the matrix of individual *z* scores on the several tests, F is the factor matrix, and P is the matrix of individual *z* scores on the *factors* involved. Since we know F and S, we can find P by premultiplying both sides of Thurstone's equation by $(F'F) - {}^1F'$ [10, p. 68], from which we get:

$$P_1 = (F'F) - {}^1F'S_1 .$$

This enables us to compute factor scores for each subject on Factor I′ and II′.[3]

It is now necessary to obtain change scores based on the variation in rigidity from the first to second testing. We obtained these by first solving the equation:

---

[3] Since the above equation assumes orthogonal factors, we rerotated the intelligence factor to an orthogonal relation to the rigidity factor. This means that the *z* scores for Factor II′ are not very pure measures of intelligence, but since they are not used in the further analysis, no harm is done.

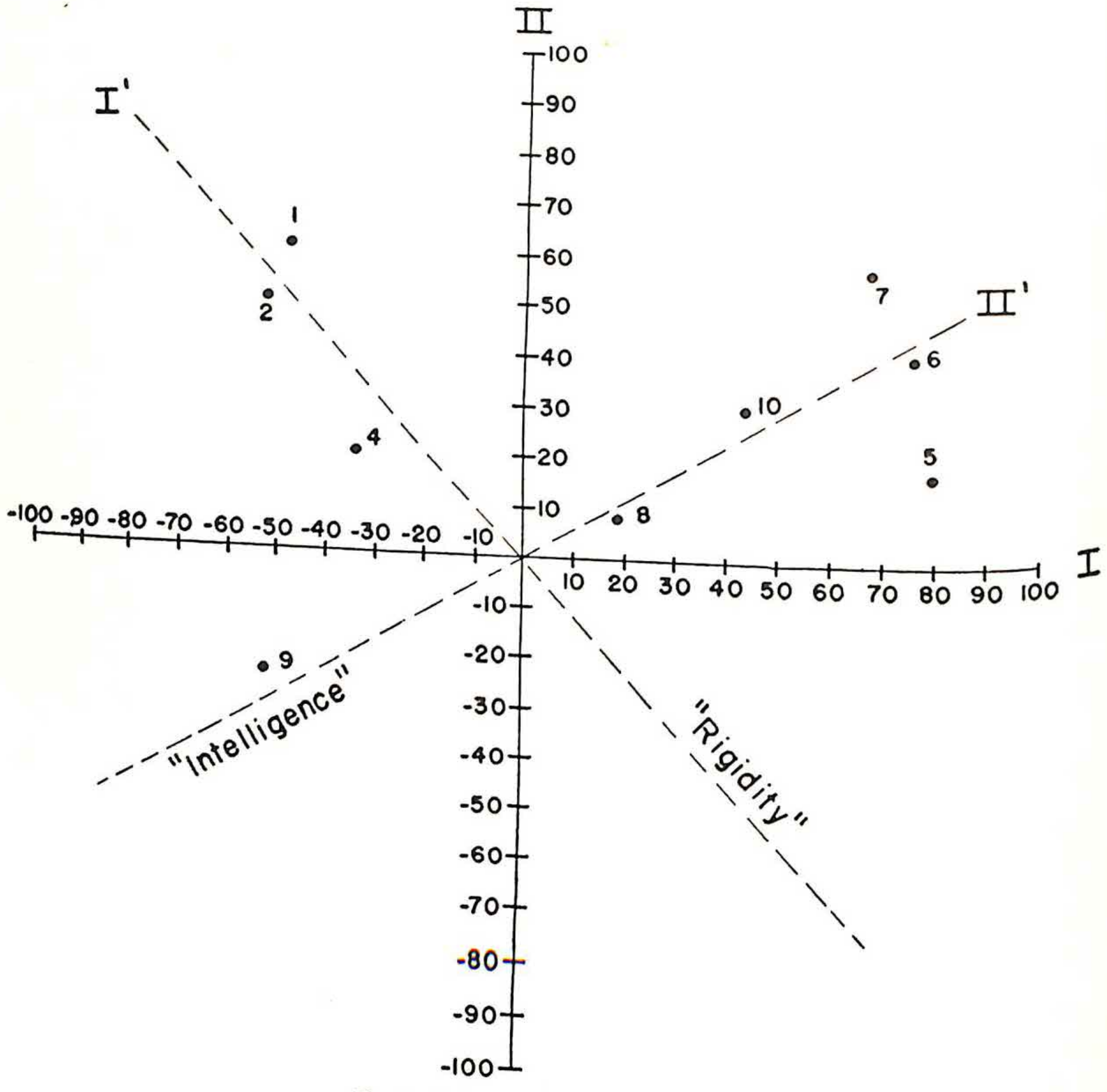


Fig 1. Original and rotated axes.

$$P_2 = (F'F) - {}^1F'S_2$$

in which the same F matrix was used, but where S_2 was computed as follows: each S's raw score on the second testing was converted into the z score he would have obtained had each distribution retained its same mean and sigma as on the first administration. We interpret these as absolute measures of rigidity based on the same scale units derived in matrix P_1.

Having determined for each S a rigidity z score for test and retest, we subtracted the first from the second to obtain a measure of change. This will be referred to as his D score. The score on the original testing will be referred to as the R score.

Results

Our principal interest is in the effect of shock therapy in reducing rigidity in the Experimental as compared with the Control group. As Table 4 shows, the mean D score for Experimentals was −.76, Controls −.53, giving a net difference in favor of the Experimentals of 0.23 z-score points. Thus both groups show a decline in rigidity, the Experimentals decreasing more. The confidence level for this difference is between .15 and .20; in other words, we cannot reject the null hypothesis and assert

Table 4

Group Comparisons of *D*-score Means*
(total sample)

Group & Mean D Score		Net Diff.	SE of Diff.	t	df	Point of Signif.†
Experimental (all) −.76	Control (all) −.53	.23	.24	.98	58	$.20>p>.15$
Experimental (improved) −.96	Control (all) −.53	.43	.24	1.83	47	$.05>p>.025$
Experimental (improved) −.96	Experimental (unimproved) −.34	.62	.45	1.36	33	$.10>p>.05$

*D scores are differences in rigidity factor scores (R scores) from Administration I to Administration II of test battery.

†One-tailed test.

that the shock therapy has significantly reduced rigidity, although the trend is in that direction.

If we reject from the Experimentals those patients rated as unimproved by the psychiatrists after shock, the mean *D* score obtained is significant at a confidence level between .025 and .05. The improved and unimproved Experimentals differ by a quantity approaching significance (confidence level between .05 and .10).

The above findings are virtually repeated if we limit our sample to schizophrenics.[4] The only interesting change in confidence levels is that the Experimental improved vs. unimproved comparison is now highly significant.

Significance of original R *score.* Despite the fact that the differences in mean *D* score consistently favor the group receiving shock, and that this difference is statistically significant for those patients rated as improved, there may be variables other than shock which could account for the finding. One of these is size of original rigidity score.

There is a correlation of −.45 between *R* score and *D* score for the total group. This means that, in general, those who were low on rigidity at the beginning have tended to increase, or those who were high to decrease in score. However, when the Experimental and Control groups are treated separately, we find that this is true only within the Experimental group ($r = -.49$), not in the Controls ($r = -.08$). Thus it cannot be simply a manifestation of the general tendency to regress toward the mean on retesting.

It is also possible to show that the differences are not due to differences between groups in original *R* score. Table 5 shows that the Experimentals and Controls differ only at the third decimal place on *R* score mean, and the Experimental Improved were actually a shade more rigid on initial testing than the unimproved fraction. None of these differences even approach significance.

Table 5

Group Comparisons of *R*-score Means*
(total sample)

Group & Mean R Score		Net Diff.	SE of Diff.	t	df	Level of Signif.†
Experimental (all) .002	Control (all) −.004	.01	.28	.02	58	$.95>p>.90$
Experimental (improved) .07	Control (all) .00	.07	.31	.23	47	$.90>p>.80$
Experimental (improved) .07	Experimental (unimproved) −.14	.21	.47	.45	.33	$.70>p>.60$

*R scores are scores on the rigidity factor on first administration of the battery.

†Two-tailed test.

Were the Controls more severely disturbed than the Experimentals prior to shock? Each patient was rated for degree of incapacitation just prior to the first testing. Table 6 shows that none of the differences between groups in degree of disturbance approached significance. It is, however, interesting to note that there was a slight (not significant) tendency for the severely disturbed group to make higher *R* scores than those less disturbed. In view of the data shown in Tables 5 and 6, it is clear that this has not contributed spuriously to the findings on the effect of shock therapy.

---

[4] This involves the elimination of only five cases. Data are not reproduced here, but can be found in full in Pullen [7].

Table 6
Distribution of "Degree of Incapacitation" (DOI) Variable

	Groups Compared		
	Experimental with control	Improved experimental with control	Improved experimental with unimproved experimental
Value of chi square	.43	1.08	1.16
Level of confidence	$.50>p>.30$	$.35>p>.25$	$.30>p>.20$

Specific tests. The data have also been compared for each of the nine tests in the final battery. Of these, tests 1, 2, and 5 (changing colors, changing figures, and hidden objects) most clearly show decrease in rigidity after shock. Tests 6, 7, and 9 (hidden words, riddles, and flicker fusion overlap) show some confirming differences. Tests 4 and 8 (motor reversal and flicker fusion threshold) contradicted the trend; shock cases made smaller changes away from rigidity than did the controls, and the Improved cases made smaller changes than the Unimproved.

Intelligence. Of incidental interest may be our finding that intelligence was not significantly affected by shock. As other studies have shown, our Experimentals showed a slight decline in Wechsler-Bellevue score, but it is particularly interesting to find that the Improved group declined less than the Unimproved shock cases. None of the differences approach significance.

Discussion

We feel that our results justify the following generalizations: (a) rigidity, defined as the tendency to persist in a previously made response, in the face of changing conditions which render the response inappropriate, is a generalized factor, manifest in a variety of different test situations; (b) rigidity, so defined, is decreased by shock therapy; and (c) persons classified as improved show more decrease in rigidity than patients classified as unimproved after shock therapy. However, we do not find, either in our own data or in the literature, a satisfactory suggestion as to the nature of the basic mechanism involved.

Because of the fairly obvious resemblance of our rigidity measures to the familiar negative transfer (proactive inhibition) experiment, an interpretation in terms of *habit interference* is inviting. The fact that our best measures, both of rigidity and of change, were perceptual in character tempts us to speculate on a parallel phenomenon of *perceptual interference*. Each such approach seems somewhat inadequate, however, because it ignores the *generalized* character of the phenomenon. We are not dealing here with a simple case in which habit X interferes with habit Y; instead, we have individual M showing, in a variety of situations, more interference than individual N.

As a tentative interpretation we offer the following speculations regarding the rigid personality. Every individual can be conceived as developing certain systems, perceptual and behavioral, which protect his essential inner equilibria [Stagner, 9]. The normal personality is characterized by perceptual systems of alertness to changing stimuli, and response sequences oriented to these external stimuli. For other personalities, alert perception of change may not have led to maintenance of equilibrium; relatively greater reinforcement may have followed from ignoring the changing stimulus, and plugging away with an invariant response. Could we describe the rigid personality as one which has learned to protect inner equilibrium by blocking out awareness of the changes in the external world? This would certainly fit some of our schizophrenics!

As soon as we start defining extremes of these response tendencies, however, we become aware that normal people show some rigid behavior, and few rigid individuals refuse completely to observe and adapt to change. The person with a clear-cut reality orientation may still, at times, withdraw and ignore the outer world, attending presumably to inner states; and correspondingly, even the withdrawn schizophrenic can be induced to respond to external stimuli if they are sufficiently demanding.

This suggests that we are dealing with a balance of two complex response systems, both being developed to some extent in every personality. Let us label as *System R* the system of perceptual and response tendencies which are closely oriented to reality. These are motive-satisfying, tension-reducing activities which involve a need to attend to changing external stimuli and to modify responses in an appropriate manner. Now let us label as *System W* the system of perceptual and response tendencies which are oriented away from the external stimulus. These likewise are activities which have served to satisfy motives, to reduce relative tension or to avoid increase of tension. Each of us feels a need, at times, to get away from painful reality. Fantasy is only one of the many devices by which we block responses to the present external world.

We propose, in brief, that all individuals develop both System *R* and System *W*, but that the *relative balance* of the two varies widely. The alert, reality-oriented extrovert has learned a predominance of *R* patterns; these have led to a maximum of reinforcement in his experience. The resistant, uncooperative, rigid schizoid has not found satisfaction in *R*-type responses; on the contrary, his motives have been channeled into withdrawing, ignoring, avoiding patterns. Obviously, under pressure the *W* system can be pushed aside and the *R* system activated; or, with the *R* system satiated, *W* responses may dominate the organism. It would be predicted that the extent of learning of such systems would determine the relative rigidity of each personality.

In the situation presented by our rigidity test battery, the patient no doubt would have preferred to continue his *W* system activities (autisms, fantasies). If, however, the *W* needs are relatively satiated, or if *R* stimuli of considerable magnitude are present, it is possible to activate the *R* system. Under pressure from the experimenter, the patient will look at the card, name the color, etc. The *R* stimulus has exceeded the threshold value needed for effective competition with the W system. However, the *R* needs satiate readily, the strong *W* needs take over, and he attempts to withdraw from the situation. If the experimenter continues to press for a response, the patient is likely to repeat his preceding response; this is available with little energy required, and it serves as a tension reducer by getting rid of the pressure from the experimenter. Thus, the schizophrenic tests "rigid" not because he is incapable of perceiving the stimulus and changing his response, but because he has adopted a different device for maintaining equilibrium. As the situation looks to him, the easiest way to keep the experimenter at arm's length and thus continue with preferred autistic responses, is to keep giving a response once found acceptable.

Why, then, does the schizophrenic change his response at all? Eventually, we assume, the growing discrepancy between the stimulus he is receiving and the response he makes reaches a threshold value; i.e., while he ignores a minor degree of "inappropriateness," a major discrepancy is potent enough to block out the *W* system and reinstate the *R* system. As the experiments on distraction have shown, we shut out disturbing external stimuli by channeling more energy into the on-going response system. However, even a highly motivated habit pattern can be disrupted if the external stimulus becomes sufficiently intense.

Does such an interpretation of personality rigidity help in the understanding of shock therapy? We believe that it does. Here we rely particularly upon the amnesic effects of shock. The *W* system of the schizophrenic is composed of percepts, autistic habits, fantasies, and avoidance behaviors of fairly recent origin. His *R* system, of observing external events, obeying verbal instructions, and responding to changes in the field, was laid down earlier in childhood. Shock, by inducing amnesia for the stronger but later habits, "uncovers" the *R* system and frees it from interference. Further, by disorganizing the fantasy world, shock may make it less attractive, so that potential gratifications in the real world seem more desirable.

Essentially, then, we are suggesting that shock therapy modifies rigidity by its effects on memory. In a more precise way, we would say that rigidity is a complex form of perceptual interference and habit interference, based upon generalized response systems. Learned motive patterns support one system oriented to reality, to changing stimuli, to demands of society, and another system involving responses of withdrawal, fantasy and avoidance of social requirements. Rigidity develops as the latter responses increasingly interfere with the former; and rigidity decreases when an event, such as shock therapy, breaks up the withdrawal habit patterns and thus eliminates interferences with earlier, reality-oriented responses.

Summary

1. We have defined rigidity as the tendency

to persist in a previously made response, in lieu of switching to one which is more appropriate to the stimulus situation.

2. Factor analysis confirms the view that rigidity as defined is a generalized feature of psychotic personalities.

3. Convulsive shock therapy produces a decrease in rigidity scores as defined; this decrease approaches statistical significance for the total group, and meets criteria of significance for those shock cases rated as improved.

4. The differences cannot be imputed to differences between experimental and control subjects in degree of original rigidity, or in degree of disturbance prior to treatment.

5. An interpretation is offered in terms of motivation to maintain equilibrium by responding alertly to reality, or by withdrawing from it. The amnesia induced by shock may break up habit systems based on the withdrawal orientation, and thus eliminate habit interference with responses based on the earlier reality orientation.

Received September 8, 1952.

References

1. Billig, O., & Sullivan, D. J. Therapeutic value of protracted insulin shock. *Psychiat. Quart.*, 1942, **16**, 549–564.
2. Cattell, R. B. The riddle of perseveration. I. "Creative effort" and disposition rigidity. II. Solution in terms of personality structure. *J. Pers.*, 1946, **14**, 229–267.
3. Cattell, R. B., & Tiner, L. G. The varieties of structural rigidity. *J. Pers.*, 1949, **17**, 321–341.
4. Fisher, S. An overview of trends in research dealing with personality rigidity. *J. Pers.* 1949, **17**, 342–351.
5. Leeper, R. W. Study of a neglected portion of the field of learning — the development of sensory organization. *J. genet. Psychol.*, 1935, **46**, 41–75.
6. Luchins, A. S. Rigidity and ethnocentrism: a critique. *J. Pers.*, 1949, **17**, 449–466.
7. Pullen, M. S. Rigidity in adult male psychotics and its modification by convulsive shock therapy. Unpublished doctor's thesis, Univer. of Illinois, 1952.
8. Rokeach, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259–278.
9. Stagner, R. Homeostasis as a unifying concept in personality theory. *Psychol. Rev.*, 1951, **58**, 5–17.
10. Thurstone, L. L. *Multiple-factor analysis.* Chicago: Univer. of Chicago Press, 1947.

Journal of Consulting Psychology
Vol. 17, No. 2, 1953

A Specific Relapse Phenomenon During the Course of Electric Convulsive Therapy[1]

Roderick W. Pugh
Veterans Administration Hospital, Hines, Illinois

This article is the report of one group of findings resulting from an exploratory, descriptive investigation of certain psychological processes during courses of concurrent electric convulsive therapy and nondirective psychotherapy with paranoid schizophrenia [1]. These findings pertain to a specifically localized relapse in charted improvement, followed by recovery, which occurred during the courses of treatment.

Problem and Procedure

The problem was to chart the course of certain psychological processes, and to determine the significance of any measurable changes in them throughout the treatment periods. The psychological processes involved were defined by eighteen categories of verbal expressions. Fourteen of these consisted of expression of attitudes or feelings and were designated "affective" categories. Four categories were positive, negative, objective, and ambivalent attitudes towards the self; four were positive, negative, objective, and ambivalent attitudes towards others; two were positive and negative attitudes perceived as directed towards the self from others; and four consisted of the summation of positive, negative, objective, and ambivalent attitudes towards the self and others. The four remaining categories of the eighteen were designated "nonaffective." They were: "Statement of problem or symptom," "Understanding or insight," "Discussion of plans," and "Irrational statements or verbalization."

The investigative procedure called for a patient's receiving sixteen electric convulsive treatments at the rate of three per week, and ten nondirective interviews approximately one hour in length. The convulsive treatments and interviews ran concurrently. The sequence began with an interview, and thereafter an interview was held twenty-four hours or later after every second treatment until the sixteen convulsive treatments and nine interviews were completed. At this point two weeks were allowed for recovery from possible "organic" effects of the convulsive therapy. Then the tenth and final interview was held. The median number of days required for the entire procedure was 56.5.

The interviews were electrically recorded and later transcribed into verbatim typescript. They were then analyzed on the basis of the eighteen categories of verbal expressions listed above so that trends followed by the psychological processes represented could be determined. This method of interview analysis yielded reliability coefficients between analysts of .88, significant at the .001 level of confidence.[2] The significance of variations in trends was derived by the statistical method of determining the significance of the difference between proportions as applied to successive proportions of interviews represented by a particular category. This same procedure was applied to groups of interviews representing the initial (interviews 1 to 3), middle (interviews 4 to 7), and final (interviews 8 to 10) stages of the treatment procedure.

---

[1] From VA Hospital, Hines, Illinois. Reviewed in the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the author are the result of his own study and do not necessarily reflect the opinion or policy of the Veterans Administration.

[2] Additional measures of over-all change were obtained by comparing before and after test data from the Rorschach, MMPI, and Stein Sentence Completion Test. These results do not feature in the findings under consideration here, however.

Subjects

The subjects were six World War II veterans drawn three each from two Veterans Administration hospitals in the Chicago area. They were selected by the criteria of a diagnosis of paranoid schizophrenia by two psychiatrists, an achievement of a *T* score of 59 or above on both the Paranoia and Schizophrenia scales of the Minnesota Multiphasic Personality Inventory, and the selection of electric convulsive therapy as a somatic treatment. They ranged in age from 24 to 35 years. The duration of their illness could not be precisely determined, but since all had served actively in the last world war, it is probable that none had been acutely affected for longer than five years prior to study. All were receiving their first course of electric convulsive therapy and none had previous exposure to psychotherapy.

The Relapse Phenomenon

Trends followed by the psychological processes in each patient over the entire treatment course predominately represented therapeutic gain. Seventy to 100 per cent of the statistically significant[3] changes recorded for each patient were so characterized. For the six patients as a group, 61 per cent of the categories (11 of 18) demonstrated significant and therapeutically directed change. The remainder showed no statistically significant changes. Greatest therapeutic gains were represented by decreased statements of problems or symptoms, and decrement in negative feelings and attitudes, especially of total negative feelings expressed and negative feelings towards the self and towards others.

During the course of treatment, however, reversals in trends indicative of negative therapeutic movement, and constituting a relapse in a general course of improvement, were ob-

---

[3] Criterion for significance is at the .05 level or less.

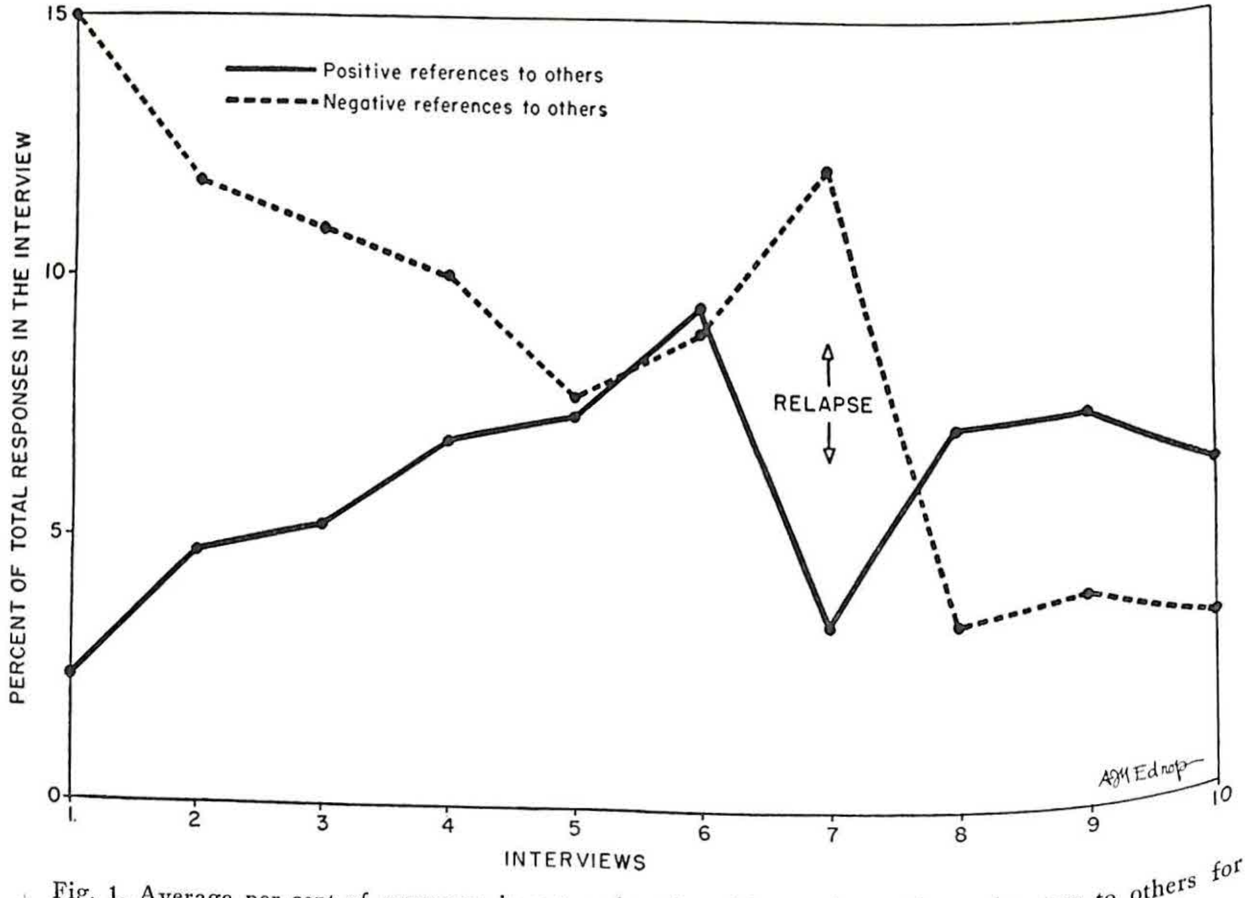


Fig. 1. Average per cent of responses in categories of positive and negative references to others for each interview for the six subjects as a group.

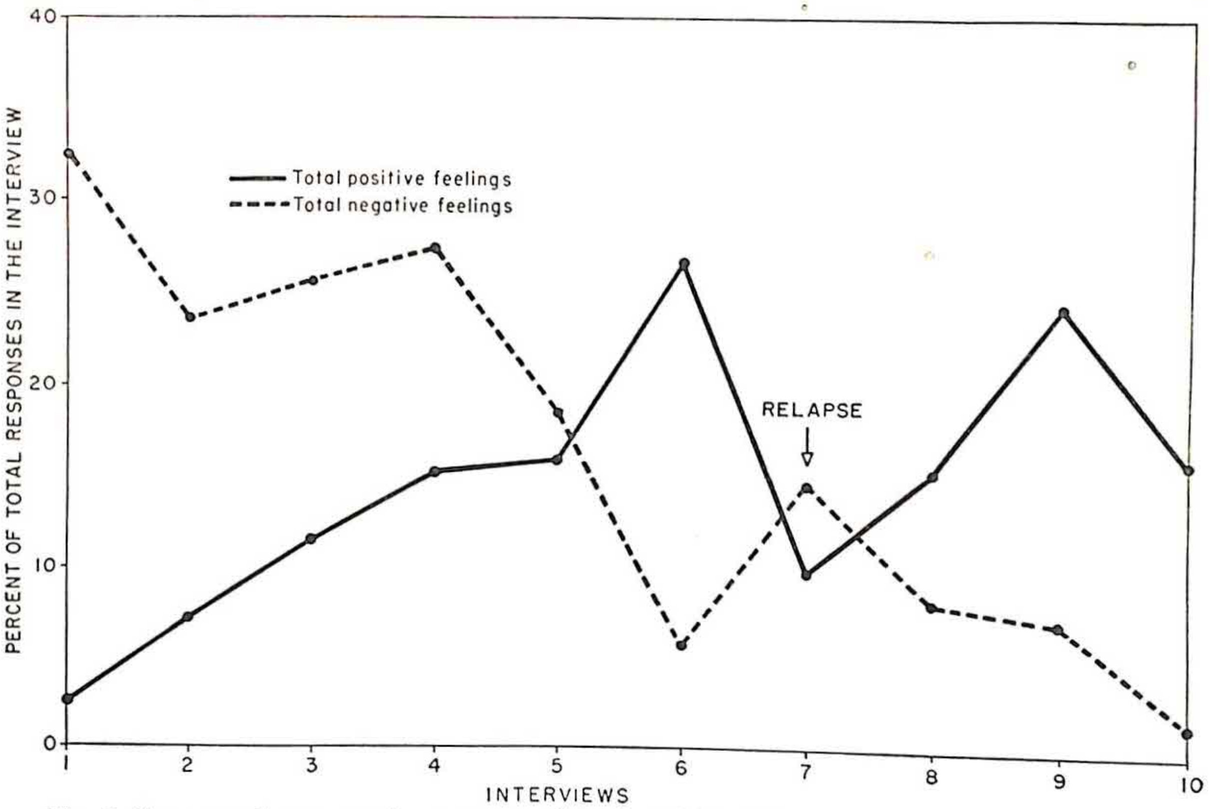


Fig. 2. Per cent of responses in categories of total positive feelings and total negative feelings expressed in each interview by Subject 4.

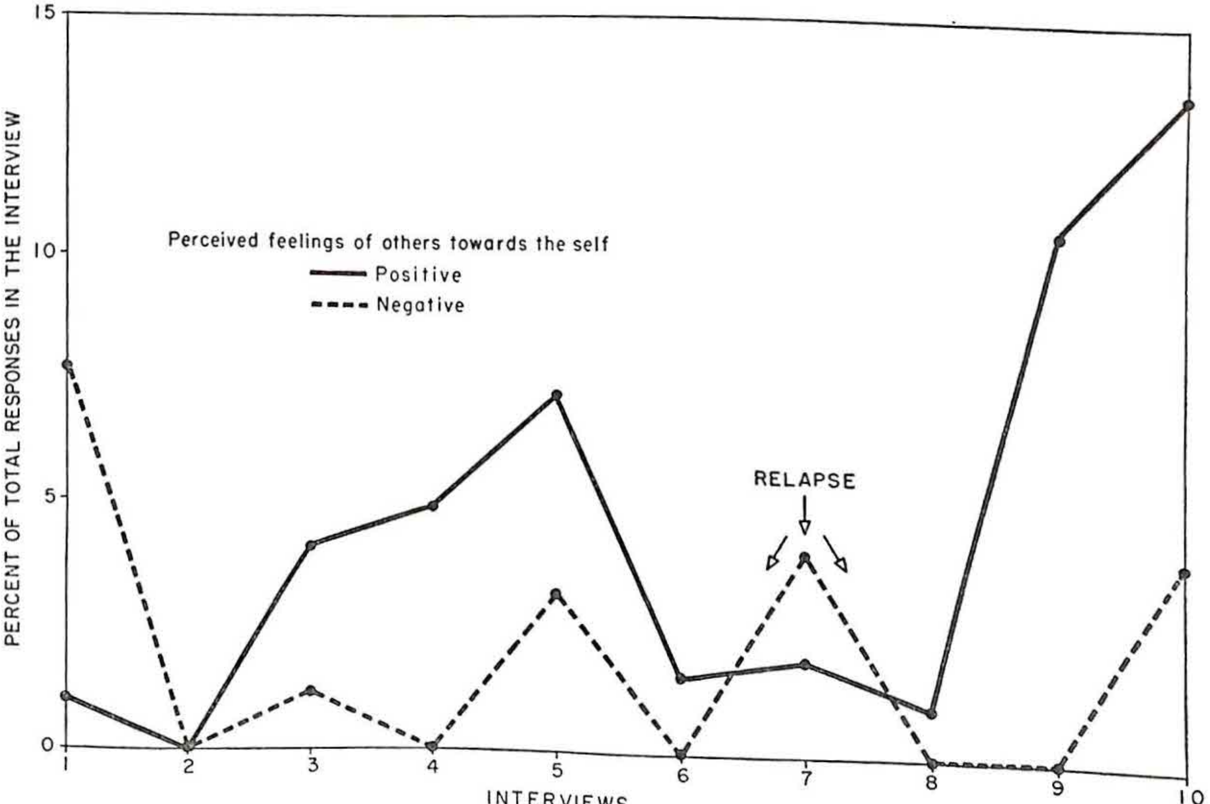


Fig. 3. Per cent of responses in categories of "perceived feelings of others toward the self" expressed in each interview by Subject 1.

served to occur specifically between the tenth and twelfth electric convulsive treatments and during the period covered by the sixth and seventh therapeutic interviews. The observed reactions were most characteristic of the affective categories, with negative feelings increased and positive feelings decreased. The strongest and most consistent reversals occurred in "Negative references to others" and in total negative feelings. Of almost equal strength and consistency were reversals in "Positive references to others" and in total positive feelings (Fig. 1, 2, and 3). Occasionally the reaction was demonstrated by decreased ability to make objective appraisal and by increased perception of negative attitudes directed towards the self from others. Some of the nonaffective categories in which the relapse phenomenon appeared in instances were "Statements of problems or symptoms" and "Irrational statements or verbalization." Interestingly, personal attitudes and feelings toward the self exhibited the greatest relative stability during the period. It should be pointed out here that the relapse phenomena were made prominent in graphical records because of early and almost immediate recovery of the previous therapeutic course of movement.

The relapse was marked in two of the six patients by clinically prominent changes in behavior such as increase in delusional material, acting-out of hostility and aggression towards others, violent rejection of the somatic therapy, and in one case, increased agitation to the point of requiring sedation. It was such a change in behavior on the part of one of these patients in particular which caused the investigator to examine more closely the therapeutic sessions of the others for the same period of the treatment course. In the case of four of the six patients, the relapse reaction was made evident only by the precise analysis of the recorded interviews. More specifically, the relapse in these four was not observed at the clinical level.

Discussion

The fact that psychiatric treatment is characterized by exacerbations, whether the treatment is somatic or psychotherapeutic, is not new knowledge. The curves of all 18 categories for each of these patients were irregular at other periods as well as at the period in question. The importance of the present observations lies in the fact that a qualitatively similar and a particularly prominent relapse, as revealed by the methods of this investigation, fell at a specific locus in treatment time for each of these six patients, in spite of the fact that the patients were not in the same phase of treatment at the same time and that they were treated at two separate hospitals. Similar data are not known to exist on any other group, nor is the investigator aware of any reported clinical observations which have so narrowly localized such a phenomenon. Each of the subjects demonstrated the relapse in from 5 to 11 of the 18 categories. Throughout the group there were 47 reversals in trends recorded during this time. Of these, 59.57 per cent were statistically significant.

Although this study was not controlled to partial out the individual effects of the convulsive therapy and the psychotherapy,[4] there seems to be little reason to doubt that the somatic treatment was largely responsible for the reactions. Since the patients were studied at different times and were divided between two different hospital environments, it is most unlikely, excluding the factor of the electric convulsive treatment, that even the same psychotherapist or that psychotherapy itself would elicit such regularly occurring reactions from six different personalities. However, it is within the realm of tenable speculation that controlled electrical stimulation to organically similar brain tissue could produce a specific physiological response from the tissue predictable with respect to time of occurrence.

The exact type of physiological reaction which might be involved is unknown. In simplest terms, however, it is possible that, while electric convulsive therapy has ameliorative effects, because of its drastic nature the attendant stress reaches a critical cumulative intensity at a certain point which is sufficiently disruptive of cerebral physiology and metabolism that prior therapeutic gains are momentarily nullified by a relapse attendant to the onset of the disruption. The organism reacts, however, by mustering all its physiological defenses to counteract this state of affairs, just as the body reacts to all states of shock. This rallying of defenses would then coincide with the recovery from the relapse. It is hypothesized that the duration of the recovery is dependent upon two factors: (*a*) the duration and intensity of continued stress stimulation, and (*b*) the inherent strength of the organism to withstand further such excesses.

---

[4] Psychotherapeutic sessions in this study were primarily intended as an avenue for investigating psychological processes rather than therapy as such. Thus, to minimize the direction of thinking and behavior, a nondirective type interview was utilized.

The findings for these six subjects are clear. The nature of the study, however, precludes broader generalizations at present. Should further research substantiate the existence of a specifically localized relapse in courses of electric convulsive therapy of ten or more treatments with paranoid schizophrenia or other diagnoses, definite implications with respect to the efficacy of longer courses of treatment are involved. Present findings suggest that future research might be directed towards investigating these additional hypotheses:

1. That the basic reaction of the relapse phenomenon is physiological.

2. That the stronger and more pervasive the relapse, the less favorable the psychological prognosis.

3. That the strength of the relapse is a positive correlate of the relative severity of the basic psychological pathology.

4. That more effective results with electric convulsive therapy are achieved if, at the rate of three treatments per week, there is sufficient interruption in time after every ninth administration to allow for physiological recovery from the stress of the therapy.

Received August 11, 1952.

Reference

1. Pugh, R. W. An investigation of some psychological processes accompanying concurrent electric convulsive therapy and nondirective psychotherapy with paranoid schizophrenia. Unpublished doctor's dissertation, Univer. of Chicago, 1949.

Journal of Consulting Psychology
Vol. 17, No. 2, 1953

Psychological Changes Following Operations on the Human Frontal Lobe[1]

Sidney Crown
Institute of Psychiatry (Maudsley Hospital), London

In a previous paper, Crown [2] reviewed studies of psychological changes following prefrontal leucotomy, the majority of which had appeared before 1949. The clinical follow-up studies discussed in that review almost invariably reported personality changes to follow the operation. These changes, however, were usually expressed in terms which are difficult or impossible to define and measure such as apathy, lack of spontaneity, fatuous equanimity, absence of finer emotional response, etc.

In order to clarify the field, it was essential that an attempt should be made to express the changes taking place after operation in terms of carefully defined, quantitative hypotheses which could be clearly confirmed or rejected. In the present state of knowledge, such hypotheses would inevitably ignore the complexity and subtlety of the changes which take place within any *individual* patient, for it would be necessary to concentrate, at first, on broad changes detectable over a *group* of persons. As in all scientific work, however, this initial oversimplification is justifiable if it leads at a later stage to more detailed experimental work.

Eysenck [5] has attempted experimentally to establish some of the dimensions or directions along which human personalities vary, and to define these dimensions operationally in terms of psychological tests. Further, in the light of reported clinical findings his conceptual scheme seemed relevant to the study of leucotomy. Studies in which an attempt was made to apply Eysenck's theory of personality organization to the neurosurgical field were, therefore, carried out by Petrie, whose reports were summarized recently by her [13], and by the present author [4].

Petrie has worked primarily with neurotic patients. She demonstrated [11] that in neurotics the personality changes three months after operation take place along three broad dimensions: in the intellectual sphere, there was a decrease in intelligence as shown particularly on the Wechsler verbal scale and the Porteus Maze test. Of the nonintellectual personality changes, there was a decrease in the conative or characterological factor that Eysenck [5] has called "neuroticism" or emotional stability; this was shown by a significant decrease in body-sway suggestibility and supported by changes on certain other objective tests and indices. The third main change was a temperamental one, from the introvert to the extravert end of the introvert-extravert continuum. This temperamental factor is one which Eysenck [5] defines in terms of tests which differentiate the dysthymic group of patients (anxious, depressed, obsessional) from the hysterics. In Petrie's study the shift on this factor after leucotomy was shown by changes such as loss of persistence in a situation involving an uncomfortable physical position, a tendency on several tests to go for speed rather than accuracy, a lowering of the level of aspiration, more realistic judgment of performance, and reduced motor perseveration. She confirmed these findings in a retest nine months after operation [10], and in a later paper [12] showed that after less severe operations the nonintellectual changes were in the same direction but were less pronounced, and changes in intelligence were not detectable at all.

Crown [4] investigated the same hypotheses in a group of patients the majority of whom were psychotic. There were few statistically significant changes in this group following the operation, although the direction of the changes was consistent with decreases in intelligence and neuroticism. It was not possible to demonstrate a consistent change towards extraversion in the psychotic patients.

---

[1] I am very grateful to Mrs. Sybil Eysenck who carried out a considerable part of the testing for this research and to Dr. H. J. Eysenck for reading and criticizing the manuscript.

That the hypotheses were clearly confirmed on neurotic and not psychotic patients is understandable in the light of later research on the "psychoticism" factor. In a recent study, Eysenck [6] confirmed the hypothesis that the functional psychoses (schizophrenia and manic depressive insanity) are not qualitatively different from normal mental states but form one extreme of a continuum which goes all the way from the perfectly normal, rational, to the completely insane, psychotic individual. This personality dimension, "psychoticism," is relatively uncorrelated with neuroticism or with introversion-extraversion. Psychotic patients, such as those tested by Crown [4], are not persons who are especially high in their position either on the neuroticism or the introversion dimensions and would not be expected to show changes along these dimensions comparable in size to those made by neurotics. Further, as has been pointed out [7], cognitive processes have not yet been linked up properly with psychoticism so that it is difficult to predict changes in intelligence for psychotic subjects after leucotomy on the basis of Eysenck's dimensional system. In the psychotic field, as Eysenck [7] further suggests, the main prediction would be that after leucotomy there would be a shift along the psychoticism axis in the direction of greater normality. A test of this hypothesis is the main interest of the present investigation.

More tests of the "psychoticism" factor than of the other factors are included in the present battery. On the factor of introversion-extraversion, on which little work has as yet been completed following up that reported by Eysenck [5],[2] only a "marker" test (persistence) is included to demonstrate the shift on this factor. On the whole, tests have been included in the present battery only if enough knowledge has been accumulated about them to place them fairly accurately as measures of a given factor. Experimental tests and tests whose position is, as yet, ambiguous because of equal saturations with more than one factor (e.g., Abstractions, cf. Eysenck [7]) have not been used.

Unfortunately, the present group contains both psychotic and neurotic patients. Also both full and limited operations have been performed. From the considerations cited earlier in this section then, it may be suggested that although the decreases in intelligence, neuroticism, introversion, and psychoticism should take place, these changes may be somewhat attenuated. If, however, the personalities of persons after prefrontal leucotomies or ablations can be shown to change measurably along precisely defined psychological dimensions, a useful preliminary aim will have been accomplished. The time will then be appropriate to attempt specific experimental studies in the neurosurgical field.

Organization of the Research

When the research was begun, it was anticipated that the patients to be operated upon would fall into three broad clinical groups: a group operated upon for the relief of obsessional illness, tension, agitation and depression, a schizophrenic group, and a group of patients suffering from intractable pain. Three types of cerebral operation were to be carried out: open prefrontal leucotomy [15], topectomy [14] and cortical undercutting [18]. As there is no adequate comparative evidence on the merits of these different types of operation, if, after careful consideration, it was considered that a patient may benefit from a cerebral operation, this was assigned to him from the three operations under investigation on a random basis. The system of randomization was such that each operation would be assigned an equal number of times within each broad clinical group. Patients considered by hospital physicians as possibly suitable for neurosurgical treatment were presented at a clinical conference and the decision for or against operation made by two senior psychiatrists and by the surgeon. In almost all cases this decision was made by the same set of persons.

---

[2] For the reader interested in the nature of introversion-extraversion, it may be stated that several follow-up researches are at present being carried out in this department.

At the present time (18 months after the start of the project) 15 patients have been operated upon and 13 of these have been followed up for at least 3 months. From the psychologist's point of view, of these 13 patients, 4 were untestable: 3 were too ill to cooperate and 1 took an early postoperative discharge. Of the remaining 9 patients, 2 could not be tested on a large number of the present tests, as these patients had been tested earlier in their stay in hospital on tests which were so similar as to invalidate testing on the present battery.

This report, therefore, is concerned with 7 patients who were tested before, and approximately 3 months after, operation. Six of the patients were woman. The age range of the whole group was from 29 years to 53 years with a mean age of 40 years. The IQ range of the 6 patients who successfully completed the Progressive Matrices test (Raven) was from under 89 to 126+ with a mean of 108 points (approximately).[3] Three of the group were suffering from severe obsessive-compulsive illnesses and four from severe depressive illnesses with marked tension. Three and possibly four of these patients were suffering from illnesses of a degree of severity which must, clinically and socially, be considered as "psychotic." Three, the obsessive-compulsive illnesses, following normal clinical practice, must be considered as "neurotic."

Three of the seven patients had an open prefrontal leucotomy, three a topectomy, and one a cortical undercutting. This unequal distribution of the various types of operation despite the over-all system of assigning operations on a statistically random basis is the result of selecting the seven patients in the present group, for reasons specified earlier, from the total number of patients operated upon to date. In the present group this meant that four out of the seven patients had an operation of limited severity. On the evidence of Petrie [12] the effect of the limited operation is to modify the severity of the postoperative personality changes. These should still, however, be in the hypothesized directions.

Method of Analysis

Lord's [9] modified *t* test[4] was used to test the significance of the differences between the scores on each test before and after operation. Taking each test separately the score of each person is compared before and after operation and a difference (*D*) column is calculated. The hypothesis to be tested is that the mean of the *D* column is significantly different from zero. In Lord's modified *t* test the test of significance is based on the use of the sample range rather than the sample standard deviation. This significance test is slightly less efficient than the more usual *t* test. It is, however, considerably easier to compute and, where the number of tests in an investigation is large and the number of persons tested small, is a useful significance test.

Table 1

Pre- and Postoperative Test Results: Intelligence, Neuroticism and Introversion-Extraversion

Test	*N*	*M* *Pre*	*M* *Post*	*s*	*p*
Intelligence					
1. Porteus Mazes	6	93.3	90.5	30.0	NS
2. Shipley-Hartford Abstractions	7	8.6	9.3	3.0	NS
3. Mill Hill Vocabulary	5	18.6	18.0	5.6	NS
4. Opposites:					
Selection	5	16.6	18.0	1.7	NS
Creation	5	16.2	14.8	0.9	0.01
Neuroticism					
5. Word Connexion List	7	13.3	12.3	7.0	NS
6. Maudsley Medical Questionnaire	7	19.9	14.4	8.1	NS
7. Perseveration	4	1.77	1.54	0.7	NS
8. Body Sway:					
Static ataxia	6	38.8	51.0	53.7	NS
Suggestibility	5	+63.2	+193.0	516.3	NS
Introversion-Extraversion					
9. Leg persistence	7	148.0	92.3	158.3	NS

---

[3] This figure is approximate because the most and the least intelligent members of the group were outside the test norms.

[4] I am indebted to Mr. A. S. C. Ehrenberg, of this Institute, for bringing the method to my attention and for his advice on other statistical aspects of this paper.

Results

The test results are given in Tables 1 and 2. In each of these tables the following legend applies: $N =$ the number of pairs of scores;[5] *M Pre* = preoperative mean score; *M Post* = postoperative mean score; $s =$ an estimate of the standard deviation of the D column (both this estimate and the tests of significance are based on the sample range instead of the sample standard deviation, cf. [9]); $p =$ level of significance (as a definite hypothesis was investigated in every case, the one-tailed test was used, cf. Jones [8]).

In the paragraphs below are given brief descriptions of the tests and methods of scoring. The numbers of the paragraphs correspond to those in the first column of the tables. Further details of these or similar procedures may be obtained from the references quoted or directly from the present writer.

Annotation to Table 1.

1. *Porteus Mazes* [16]. Procedure as detailed by Porteus. Score: Test Quotient.

2. *Shipley-Hartford Abstractions* [19]. A short reasoning test, the items of which are steeply graded in difficulty. Score: number of items correct.

3. *Mill Hill Vocabulary* [17]. A vocabulary test of the synonyms type. Alternate forms given before and after operation. Score: number of items correct.

4. *Opposites* [1]. The two alternate forms of Cattell's opposites tests were used, the first form to be administered in the usual way and then the alternate form to be administered without giving the *S* the usual choice of answers but making him supply his own answers without help. The first test is intended as a "selection" test, the second as a "creation" test.[6] Both tests were administered without time limit. Scores: number of items correct in each case.

5. *Word Connexion List* [3]. A controlled association test in which each item consists of a stimulus word and two possible response words, one of which is an "abnormal" response, the other a "normal" response. Score: number of "abnormal" choices underlined.

6. *Maudsley Medical Questionnaire* [5]. This test consists of questions dealing largely with physical complaints such as are frequently made by neurotic subjects; a few psychological questions are also included. Score: number of symptoms shown.

7. *Perseveration* [6]. The *S* is required to write a series of *S*s for 15 seconds (Period A), then a series of *S*s reversed for 15 seconds (Period B) fol-

Table 2

Pre- and Postoperative Test Results: Psychoticism

Test	N	*M Pre*	*M Post*	s	p
1. Fluency:					
Birds	7	9.7	9.9	1.5	NS
Animals	7	10.4	13.1	1.5	<0.01
Flowers	7	12.6	14.4	3.3	NS
Total	7	32.7	37.4	4.4	0.02
2. Mirror Drawing:					
1st Trial	7	165.7	80.6	66.2	0.01
2nd Trial	7	88.3	52.4	44.7	0.05
3rd Trial	7	65.6	41.1	45.9	NS
3. Circles:					
Time	7	20.3	20.4	20.0	NS
Largest Diameter	7	40.6	42.7	11.1	NS
Smallest Diameter	7	35.0	36.0	15.5	NS
Average Diameter	7	37.9	39.3	10.4	NS
4. Squares:					
Time	7	30.1	23.9	26.6	NS
Largest Diagonal	7	45.3	47.3	14.8	NS
Smallest Diagonal	7	41.1	38.4	10.4	NS
Average Diagonal	7	43.4	42.9	9.6	NS
5. Reading Prose	7	22.9	21.3	3.0	NS
6. Estimating distances:					
Overestimation 3 tests	7	2.6	1.9	6.3	NS
Underestimation 3 tests	7	21.1	13.0	15.5	NS
Error 8″ estimate	7	4.7	2.7	4.4	NS
Error 1′ estimate	7	6.4	4.0	5.9	NS
Error 2′ estimate	7	12.6	8.1	6.7	NS
Total error	7	23.7	14.9	15.2	NS
7. Waves:					
Amplitude	4	24.5	22.5	2.4	NS
Wavelength	4	116.2	90.0	15.5	0.02
8. Tapping	7	57.6	66.2	16.6	NS
9. Reversal of perspective:					
Passive	6	3.3	4.3	2.2	NS
Active	6	8.2	6.8	7.5	NS
Inhibited	6	2.5	3.0	2.8	NS

---

[5] For practical reasons it was not always possible to test or to score all the subjects on every test.

[6] This principle is a fairly common one in the design of high-level intelligence tests. The particular test used in the present investigation is based on a suggestion of Dr. H. J. Eysenck.

lowed by a 30-second period (Period C) in which he writes alternately one *S* followed by an *S* reversed. Score: $A + B/C$.

8. *Body Sway* [5]. From a string attached to the collar a continuous record is made on a kymograph of the *S*'s forward and backward sway during (*a*) a 30-second period when he is told simply to stand quite still with his eyes closed (static ataxia); and (*b*) a 2½-minute period when he is given continuous suggestion by gramophone record that he is falling forward. The kymograph records are scored graphically using a grill of ⅛-inch squares drawn on tracing paper to estimate the amount of movement, i.e., the area under the curves. Scores: Static ataxia: total forward sway (number of ⅛-inch squares) + backward sway. Suggestibility: total forward sway or total backward sway whichever is the greater.[7]

9. *Leg persistence* [5]. The *S*, sitting, holds his leg out horizontally over a chair of the same height for as long as he can without touching down. Score: time in seconds.

Annotation to Table 2.

The reference for all the tests of Table 2 is Eysenck [6].

1. *Fluency.* Number of birds, animals, and flowers the *S* can name, one minute for each.

2. *Mirror Drawing.* The *S* is required, looking only in the mirror, to trace round a diamond shape the corners of which are indicated on the paper by dots. Three trials. Score: time (seconds).

3. *Circles.* The *S* was handed a clean piece of paper and asked to draw three circles. Scoring as in Table 2: time in seconds, measures in mm.

4. *Squares.* The paper with the circles on was reversed and the *S* asked to draw three squares on the other side. Scoring as in Table 2: time in seconds, measures in mm.

5. *Reading Prose.* The *S* was asked to read aloud a standard piece of prose. Score: time in seconds.

6. *Estimating Distances.* The *S* is asked to indicate what he considered 2 feet distance was by placing two matches two feet apart on the table; then one foot; then 8 inches, so that there are four matches on the table. These distances are measured and recorded. Scoring as in Table 2: quarter-inch units.

7. *Waves.* A sheet of foolscap size, squared paper with 4 *V*'s marked at precise places is given to the *S*. In each of 4 trials the *S* is required, first, with the eyes open, to trace over a *V*; then, with eyes closed, to make six more *V*'s along the same line and, as far as possible, the same size. The arm is kept up all the while, only the pencil touching the paper. Two scores were derived: amplitude (average of 1st and last amplitude of 4 sets of waves, i.e., of 8 measures altogether); and wavelength (average wavelength of all 4 sets measured parallel to edge of the paper). Measure in mm.

8. *Tapping.* Tapping with a pencil on a sheet of paper for two trials of 15 seconds each. Score: Average number of taps made. Instruction is *not* to tap as fast as possible; score, therefore, represents natural tempo rather than maximum rate.

9. *Reversal of perspective.* Necker Cube. Illusion explained to *S*. Scores: (*a*) passive, i.e., average number of fluctuations during two 30 second periods, the one period at the beginning and the other at the end of the test. (*b*) Active, i.e., number of fluctuations recorded for 30 seconds when *S* is asked to make illusion change as fast as possible. (*c*) Inhibited, i.e., number of changes recorded during 30 seconds when *S* is asked to try and prevent the cube from changing.

Discussion

As the hypotheses investigated are precise, in the following discussion importance is attached both to changes which are statistically significant at or less than the 10 per cent level and to groups of changes which, although individually insignificant, are in the hypothesized direction.

Considering first changes in intelligence, 5 scores were derived from the intelligence tests. Of these, the changes in 4 are insignificant and their direction is inconsistent: performance is improved slightly postoperatively in the Shipley-Hartford Abstractions test and in the selection of opposites and is diminished on the Porteus Mazes and the selection of synonyms (Mill Hill Vocabulary).

There is, however, one particularly interesting change on the "Creation" test, one of the few experimental tests in the battery included in an attempt to measure changes in some of the higher intellectual functions. After a cerebral operation the group were affected to a significant degree in their ability to think out without help a word opposite in meaning to a stimulus word, although on the

[7] We have not used the kymographic method of recording body sway before and the scoring problems raised by this method have not yet been systematically tackled. In particular, falls raise difficulties of scoring and the records of persons who fell have not been included in the present quantitative analysis. However, of the falls which occurred, 3 took place before operation and one postoperatively. This is consistent with a decrease in neuroticism. Secondly, the quantitative scoring units (in this study ⅛-inch squares) are arbitrary units. It is hoped to develop a more satisfactory scoring system when the precise amount of bodily movement represented by a given amount of kymographic movement has been investigated.

"Selection" test, using words of a parallel degree of difficulty, they show themselves to be unaffected or even improved slightly in the power to select the correct opposite from a number of choices. Not enough persons were examined to make a serious quantitative analysis of errors worth while. However, considering the records of the 3 persons in the group whose scores on the "Creation" test dropped most postoperatively, the errors appeared to be of two types. In the first type an opposite of inferior stringency was given after operation compared with that given before, e.g., patient A.M. "Brave" is the opposite of "cowardly" (preoperatively), "stupid" (postoperatively); patient A.W.S. "all" is the opposite of "none" (preoperatively), "a few" (postoperatively). The second type of inferior response occurred when the *S* made no real effort to find an opposite, e.g., patient H.E.L. "Brave" is the opposite of "coward" (preoperatively), "not brave" (postoperatively); or "save" is the opposite of "spend" (preoperatively), "not save" (postoperatively).

Thus although the results of the usual intelligence tests—performance (Porteus), reasoning (Shipley) and verbal (Mill Hill Vocabulary, selection of opposites)—showed no impairment, the results from the "Creation" test do suggest that high-level intellectual changes may take place after neurosurgical operations.

Petrie [12], using the Wechsler-Bellevue vocabulary and other verbal tests, has also reported that her patients show carelessness in the use of language after leucotomy and she suggests that these intellectual changes may be more characteristic of the posterior operations. In the present small series of 5 cases tested on this test, of three patients who showed the greatest postoperative deficit on the creation of opposites test, two had the open prefrontal leucotomy, the most severe of the operations carried out in our series.

It is difficult to determine how far such changes are a "true" intellectual deficit and how far they are secondary to changes of a nonintellectual type such as in the will or motivation to work or carelessness at work. This is the difficulty which Petrie [13] also came up against in attempting to explain her findings. She provides no convincing evidence for deciding whether the change is a "true" intellectual change or not, and in the present experiment it is equally impossible to do this.

On the tests of neuroticism, none of the changes is significant but the majority are in the hypothesized direction. The group score less abnormally on the Word Connexion List, on the Maudsley Medical Questionnaire, and lower on the test of Perseveration. Only on the Body Sway test are the scores in the wrong direction for the hypothesis. There is, however, considerable variability in individual performance on both the tests of Static Ataxia and Suggestibility after operation as the statistics quoted in the previous section show. Larger numbers are needed to determine with any confidence whether the changes in score on this test are truly inconsistent or whether a more definite change in one direction or another may be discerned. Further, as stated earlier, falls were not taken into account in the present scoring although the greater preoperative occurrence of this phenomenon is as expected. On the whole, then, the results show a discriminable but attenuated shift in the scores of the group towards the normal end of the normality-neuroticism continuum. On a mixed group of neurotic and psychotic patients this result is what, as stated earlier, would be expected.

A single "marker" test, Leg Persistence, was included to check the tendency of the group to move towards the extraverted end of the introversion-extraversion personality dimension. As with the tests of neuroticism, only a slight decrease in score on this test would be hypothesized in a mixed neurotic and psychotic group. Thus the observed, but nonsignificant, decrease in leg persistence is as expected.

Of the 28 scores derived from the tests of the psychoticism factor, 21 are in the direction which would be hypothesized if the scores of the group are becoming more typical of the normal than the psychotic person, and of these 5 are statistically significant. The simple binomial test shows that the occurrence of 21 out of 28 changes in the hypothesized direction is significant (i.e., would not be accounted for by chance) assuming that each of the scores is independent. This is not so, however, as several scores are often derived from the same test. Accordingly, this significance test must be interpreted with caution. If it had been insignificant, the data would hardly have been worth considering further. Being significant, but under these limiting conditions, the results may be regarded as at least suggestive.

If Eysenck's [6] paper, in which he has demonstrated some of the personality traits which distinguish the normal from the psychotic person at a high statistical level of confidence, is taken as the basis for our hypotheses, then the shift towards normality on the psychoticism factor is shown clearly on all the tests included in the present battery except the Drawing Circles test. The shift is shown especially well on the Fluency tests, the Mirror Drawing test, and the Waves test. Thus, after leucotomy the group becomes more fluent, performs better on the Mirror Drawing test, increases the rate of oscillation on the Reversal of Perspective test, makes smaller movements on the Squares and Waves tests, shows a lesser tendency to overestimate distances and reads and taps more quickly. Despite the small numbers and the fact that the group includes both neurotic and psychotic patients, these changes support the hypothesis of a decrease on the psychoticism factor fairly convincingly.

On the whole, therefore, the data confirm the hypotheses. With a mixed neurotic and psychotic group, where operations both of limited and greater severity have been carried out, attenuated but measurable psychological changes take place postoperatively. At a high level, intellectual deficits may occur. In addition there are changes on nonintellectual personality factors such as introversion-extraversion, neuroticism, and psychoticism which are measurable at least on a group basis. A detailed exploration of the factors underlying these broad personality changes should now be profitable.

Summary and Conclusions

This study is one of several in which an attempt has been made to apply Eysenck's theory of personality organization to the neurosurgical field.

Objective tests were used to measure some of the personality changes taking place after operations on the frontal lobe.

In terms of tests used to define these factors, it was hypothesized that, postoperatively, patients would tend to show a decrease in intelligence, a temperamental change from introversion towards extraversion and characterological changes towards normality on the neuroticism and psychoticism dimensions.

Despite the small numbers studied, the inclusion of both neurotic and psychotic patients, and operations of different degrees of severity, the data tend to confirm these hypotheses.

So far as this is possible, attempts should now be made to study the mechanisms underlying these observed personality changes.

Received August 11, 1952.

References

1. Cattell, R. B. *Cattell Intelligence Tests.* Scale II: Forms A & B. London: Harrap, 1935.
2. Crown, S. Psychological changes following prefrontal leucotomy; a review. *J. ment. Sci.*, 1951, **97**, 49-83.
3. Crown, S. The Word Connexion List as a diagnostic test: norms and validation. *Brit. J. Psychol.*, 1952, **43**, 103-112.
4. Crown, S. An experimental study of psychological changes following prefrontal lobotomy. *J. gen. Psychol.*, 1952, **47**, 3–41.
5. Eysenck, H. J. *Dimensions of personality.* London: Kegan Paul, 1947.
6. Eysenck, H. J. Schizothymia-Cyclothymia as a dimension of personality. II. Experimental. *J. Pers.*, 1952, **20**, 345-384.
7. Eysenck, H. J. *The scientific study of personality.* London: Routledge & Kegan Paul, 1952.
8. Jones, L. V. Tests of hypotheses: one-sided vs. two-sided alternatives. *Psychol. Bull.*, 1952, **49**, 43-46.
9. Lord, Edith. The use of range in place of standard deviation in the *t*-test. *Biometrika*, 1947, **34**, 41–67.
10. Petrie, Asenath. Personality changes after prefrontal leucotomy. *Brit. J. med. Psychol.*, 1949, **22**, 200-207.
11. Petrie, Asenath. Preliminary report of changes after prefrontal leucotomy. *J. ment. Sci.*, 1949, **95**, 449-455.
12. Petrie, Asenath. A comparison of the psychological effects of different types of operations on the frontal lobes. *J. ment. Sci.*, 1952, **98**, 326-329.
13. Petrie, Asenath. *Personality and the frontal lobes.* London: Routledge & Kegan Paul, 1952.

14. Pool, J. L. Topectomy. *Lancet,* 1949, **2,** 776-781.

15. Poppen, J. L. Technic and complications of the standard prefrontal leucotomy. In M. Greenblatt, R. Arnot, & H. C. Solomon. (Eds.), *Studies in lobotomy.* New York: Grune & Stratton, 1950.

16. Porteus, S. D. *The maze test and mental differences.* Vineland, N. J.: Smith Printing House, 1933.

17. Raven, J. G. *The Mill Hill Vocabulary Test.* London: H. K. Lewis, 1944.

18. Scoville, W. B. Selective cortical undercutting as a means of modifying and studying frontal lobe function in man. *J. Neurosurg.,* 1949, **6,** 65-73.

19. Shipley, W. C. A self-administering scale for measuring intellectual impairment and deterioration. *J. Psychol.,* 1940, **9,** 371-377.

Journal of Consulting Psychology
Vol. 17, No. 2, 1953

Generalization of Problem-Solving Rigidity[1]

Emory L. Cowen,[2] Morton Wiener, and Judith Hess
University of Rochester

Problem-solving or Einstellung rigidity has been defined as the tendency to adhere to an induced behavior when it ceases to represent the most direct path to a goal [4]. Considerable research interest has surrounded this phenomenon in recent years, both in terms of its basic parameters and alleged personality correlates. With respect to certain aspects of the problem, there appears to be considerable agreement amongst the findings of independent researchers to date. Thus for example, starting with the classic studies of Luchins [8] in 1942, the susceptibility of Einstellung rigidity to a variety of specifiable field conditions now appears to be firmly established.

Perhaps the single issue in this general area of research about which agreement is most lacking is whether or not this type of behavioral rigidity bears relationship to other personality variables. That such a relationship may exist is implied in the work of the California group [1, 6], particularly that of Rokeach [10], as well as in the studies of Cowen and Thompson [4], Fisher [5], and Harway [7]. Luchins in his critique of Rokeach's work [9] takes an opposite viewpoint stating that "rigidity is not a function of the personality per se, but of particular field conditions."

In the present experimental approach to this problem, we have started with the observation that when all pertinent variables are controlled (i.e., intelligence, age, etc.) and under a constant set of field conditions, individual differences in rigidity of response continue to be manifested. If, with the necessary controls exercised, it can be demonstrated that this state of affairs is not solely a function of specific task attitudes (i.e., toward arithmetical problems in the case of the water-jar technique), we shall be able to lend greater credence to the existence of some concept of a personality-related mode of problem-solving behavior — at least under a given set of field conditions.

The research problem, viewed in these terms, becomes a relatively simple and clear-cut one which may be described in two steps:

1. To develop additional measures of problem-solving rigidity which are structurally similar to the water-jar technique, but which tap different areas of functioning.

2. To test the hypothesis that there will be a positive relationship between rigidity behaviors as measured by the several tasks.

Procedures and Results

Subjects in the present study were 59 college undergraduates, 30 females and 29 males. Each *S* was seen individually by one of three examiners for two testing sessions approximately four weeks apart.

Three tests were administered, each designed to measure problem-solving rigidity in a different area of functioning (i.e., the Luchins water-jar technique, an original alphabet maze, and an original motor maze). These tests were constructed so as to be structurally similar and capable of measuring the tendency to adhere to an induced method of problem solving when this method ceases to be the most direct one. In the case of the motor maze, however, the selected problems

---

[1] Portions of this research were presented in a paper read to the Division of Personality and Social Psychology, American Psychological Association, Washington, D. C., September, 1952.

[2] This investigation was supported in part by a research grant to the senior author from the National Institute of Mental Health, National Institutes of Health, Public Health Service. The specific study reported in the present paper is part of a broader research program in the area of sociopsychological and personality correlates of psychological rigidity.

PRACTICE

N	E	V	E	S
F	P	M	L	X
I	G	K	Q	T
S	U	O	S	Z
H	E	Y	V	R

SOLUTION:

SEVEN FISH

SET

G	M	O	E	Y	L
Z	T	U	K	E	S
D	S	H	T	Q	E
A	V	C	J	H	F
B	W	F	Z	I	C
G	N	I	S	M	R

SOLUTION:

LET HIM SING

CONTROL-CRUCIAL

I	O	B	C	L	H
M	N	T	Q	E	Z
H	G	F	R	Z	O
D	V	H	Y	R	J
X	A	G	Q	E	V
T	I	U	S	D	W

LONG SOLUTION:

HER RED SUIT

DIRECT SOLUTION:

HER HAT

Fig. 1. Sample problems from the alphabet maze series.

proved to be too simple; hence no measure of rigidity was obtained. Primary consideration will therefore be given to the relationship between the water-jar and alphabet-maze techniques.

With respect to administration, the water-jar technique was always given first. Following this, one-half of the *S*s received the alphabet-maze technique and the remainder were given the motor mazes. The third rigidity measure was given in the second testing session. This procedure was followed to determine whether obtained correlations between rigidity measures would differ as a function of the temporal interval between tests. Since this was found not to be the case, it was considered defensible to pool alphabet-maze scores from the two administrations.

Since various adaptations of the water-jar technique as well as instructions for its administration are reported elsewhere [3, 4, 8, 10], the actual test problems used will not be reproduced here. Suffice it to mention that the series consisted of two practice problems, designed to familiarize *S*s with the various steps that could legitimately be utilized in the solution of later problems; one control problem which could be solved either by a long method (B-A-C-C) or a direct one (A-C) which is actually one of the steps of the long method; five Einstellung or set problems soluble only by the long method; four crucial problems, similar to the control problem and soluble either by the long or direct method; one extinction problem soluble only by a new and direct method not used in any previous problem; and finally two additional crucial problems.

No published account of the alphabet-maze technique is available in the psychological literature, although Cowen [2] reports early pilot work with this technique. For present experimental usage the test series was arranged so as to parallel exactly the water-jar technique. The same number of practice problems, controls, set-builders, crucials, and extinctions were used, in the same order as for the water-jar technique.

In the administration of the alphabet-maze technique, *S* is told that the object of the task is to move from the upper right-hand corner of the maze to the lower left-hand corner, spelling out words on the way. He may move one box at a time in any direction (up, down, left, right, or diagonal) as long as the move he makes helps to spell out a word. The answer which he is to write in the booklet consists of the word or words that will take him from start to finish of the maze. In case more than one path is available, the correct solution is the one that uses the fewest number of boxes. Figure 1 consists of a sample practice, set, and "control-crucial" problem used in this series.

The long solution used consists of moving along a down-left diagonal, then straight down, and finally directly to the left. The direct solution is a straight diagonal path from the upper right-hand corner to the lower left-hand corner. The long solution requires nine moves; the direct one only five moves.

Problems in both test series were presented one by one and in order, on individual three-by-five index cards. Responses were recorded by *S*s in examination booklets — a fresh page being taken for each problem.

The rigidity score for each subject on each instrument consisted of the total number of crucial problems solved by the long method. In the case of the extinction problem nonsolution in the 2½-minute time period allotted was considered to be a rigid response. Thus, the highest possible rigidity score on each instrument was seven.

Subjects were eliminated for either of two reasons: (*a*) a long solution on the control problem, and (*b*) mathematical or word errors on the test problems.

By these criteria, 47 subjects had acceptable records on both tests. The computed Pearson *r* between water-jar and alphabet-maze rigidity was .42 which was significant beyond the .01 level. Since an analysis of test scores indicated essentially dichotemous, U-shaped distributions for both measures, a phi coefficient was computed and found to be .46. The significance of this coefficient as tested by χ^2 was beyond the .01 level of probability. A tetrachoric *r* of .59 estimates the degree of relationship between the instruments based on the assumption of normality of distribution.

Discussion

General confirmation of the original hypothesis is found in the low positive but significant correlation between the two rigidity measures. Under the specific and constant field conditions used in the present investigation, and with other pertinent variables nonoperant, individuals do respond differentially in terms of rigidity behavior on any given task. Moreover, we can not account for such response tendencies solely on the basis of specific task attitudes, in the light of its generalization to a different area of functioning.

Though the present investigators attach importance to this finding, caution must be exercised in its interpretation and generalization. A sweeping generalization, such as that problem-solving rigidity is a function of personality, is not indicated or implied. Quite conceivably there could be situations in which field factors are so compelling as to preclude entirely the operation of personality factors. Illustratively, had the experimenters used 100 set problems, very likely all *S*s would have responded "rigidly" on all subsequent crucials. On the other hand there appears to exist a potential "balance point" in experimental field conditions, at which a relatively wide range of rigidity responses are elicited from individuals of comparable age, intelligence, and education. Under such conditions in the present study, a tendency towards a generalized mode of problem solution has been demonstrated, and the likelihood that this tendency is "personality-related" strengthened.

The precise relationship between rigidity and other personality variables is likely to vary considerably under different field conditions. It may well be that this fact underlies many of the discrepant findings in this area to date. We advance a general hypothesis, however, that the relationship will be greatest when field conditions are such as to bring about, in an otherwise homogeneous group of subjects, a U-shaped type distribution of rigidity scores described above.

A final issue which merits some consideration centers around the question of what is the most fruitful direction that subsequent research in this area can follow. We propose that it is not helpful and very probably not even meaningful to ask a question such as, "Is there a relationship between rigidity and personality?" Such a phrasing, while perhaps inviting to those who seek simple linear relationships, overlooks the multiplicity of settings in which the covariation of these variables may conceivably be studied. The present authors view with considerably greater optimism an approach which seeks to specify precisely a wide variety of experimental field conditions and to examine the relationship between problem-solving rigidity and other designated personality variables, under these measurable conditions.

Summary

In order to test the hypothesis of generalization of problem-solving rigidity the authors have developed an alphabet-maze scale, designed to parallel the Luchins water-jar technique in structure, but to deal with a different area of cognitive function.

The two instruments were given to 59 college undergraduates, and on the basis of the scoring criteria used, 47 pairs of usable test records were obtained. The significant correlation between the two measures was taken as a confirmation of the original hypothesis. Under the particular field conditions employed, a generalization of problem-solving rigidity has been demonstrated.

The limitations of the present findings, as well as their implications for subsequent research have been considered.

Received August 11, 1952.

References

1. Christie, J. R. The effects of frustration on rigidity in problem solution. Unpublished doctor's dissertation, Univer. of California, 1949.
2. Cowen, E. L. A study of the influence of varying degrees of psychological stress on problem solving rigidity. Unpublished doctor's dissertation, Syracuse Univer., 1950.
3. Cowen, E. L. Psychological stress and problem solving rigidity. *J. abnorm. soc. Psychol.*, 1952, **47**, 512-519.
4. Cowen, E. L., & Thompson, G. G. Problem solving rigidity and personality structure. *J. abnorm. soc. Psychol.*, 1951, **46**, 165-176.
5. Fisher, S. Patterns of personality rigidity and some of their determinants. *Psychol. Monogr.*, 1950, **64**, No. 1 (Whole No. 307).
6. Frenkel-Brunswik, Else. Intolerance of ambiguity as an emotional and perceptual variable. *J. Pers.*, 1949, **18**, 108-143.
7. Harway, N. I. Personality variables in problem solving rigidity inferred from behavior in the level of aspiration situation. Unpublished doctor's dissertation, Univer. of Rochester, 1952.
8. Luchins, A. S. Mechanization in problem solving: The effect of Einstellung. *Psychol. Monogr.*, 1942, **54**, No. 6 (Whole No. 248).
9. Luchins, A. S. Rigidity and ethnocentrism: A critique. *J. Pers.*, 1949, **17**, 449-466.
10. Rokeach, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259-278.

Journal of Consulting Psychology
Vol. 17, No. 2, 1953

The Use of Rorschach Scores to Predict Whether Patients Will Continue Psychotherapy

Frank Auld, Jr., and Leonard D. Eron[1]
Yale University

Interpreting psychological tests is a serious matter, for interpretations affect the lives of patients. Decisions to accept patients or reject them, to try psychotherapy or not, to use one method of treatment instead of another—all these may be made on the basis of psychological reports. The psychologist, then, has an ethical responsibility to be sure that his statements are founded on scientific knowledge. He cannot afford to rely on guesswork; he must be as right as scientific method can enable him to be. In this paper we point out that the clinical psychologist who wants to be right must be sure that the diagnostic devices he uses have been cross validated. Of course, practically all psychologists agree with this proposition in the abstract. We hope to buttress the logic of this matter with concrete examples so that the reader will never be able to forget the rule: *If the combining weights of a set of predictors have been determined from the statistics of one sample, the effectiveness of the predictor-composite* must *be determined on a separate, independent sample* [Mosier, 9].

Background of This Study

Many psychiatric clinics have long waiting lists and can accept for treatment only a few of the many people that apply. Thus there is a great need for diagnostic methods that can predict which of the applicants can profit most from therapy. Kotkov and Meadow [6, 7] have proposed a method of doing this by means of certain Rorschach scores. These authors tried out all of the common Rorschach scores; they found only three that discriminated between patients who continued for at least nine interviews and patients who quit. By means of Fisher's discriminant function Kotkov and Meadow combined these scores into a prediction formula: $Y = .00038\,R + .00007\,D\% + .00241(FC-CF)$. ($R$ is the total number of responses to the Rorschach; $D\%$, the percentage of these that are large usual details; $FC - CF$, the number of form-color responses minus the number of color-form responses.) The weights for this formula were derived from a sample of 98 patients who received group psychotherapy at the Boston VA Mental Hygiene Clinic. When applied to this sample, from which the weights had been derived, the formula correctly classified 69% of the subjects as "continuing" or "discontinuing" patients.

Kotkov and Meadow then applied this formula to a sample of 52 individual psychotherapy patients in the Boston clinic. Note that this is not a classical cross-validation design, since the patients were not drawn from the same population (namely, group therapy patients) as the patients in the original sample. This type of study, in which weights determined on a sample from one population are applied to a second sample drawn from a differently defined population, may be called "validity generalization" [Mosier, 9]. When applied to the individual therapy sample the formula discriminated continuers from quitters just as well as it had in the original sample of group therapy patients, making correct predictions for 36 of the 52 patients (69%). Not all of the scores entering into the formula, however, discriminated continuers and quitters among the individual psychotherapy patients;[2] only the $FC-CF$ score did so to a statistically significant degree. The R score was almost significant for the group therapy patients and statistically significant for the individual therapy patients. The $D\%$ score was significant at the .30 level for the group therapy patients, but the direction of the difference was reversed in the individual therapy sample. It is surprising that there was no loss in predictive efficiency when the formula was applied to a new sample. Shrinkage in multiple correlation (the discriminant function is analogous to multiple correlation) has been the regular finding of psychologists who have utilized prediction formulas.

---

[1] The authors wish to thank Dr. Mark A. May, director of the Institute of Human Relations, for the Institute's financial support of this study. Dr. Neal E. Miller read this paper and made helpful suggestions.

Since predictions stemming from this formula are based solely on the addition of three scores, it might be felt that the formula does not get the best possible predictions out of the data. We believe, however, that the formula probably does as well with the materials it embraces (the commonly used Rorschach scores) as any subjective method for making predicitions from these scores would do. The reasoning of Thorndike [12, pp. 200-201] and the experience of Kelly and Fiske [4, pp. 200-202] support this view.

Although Kotkov and Meadow's prediction formula worked well when applied to the Boston samples, we felt that it might not be equally effective when applied to new samples of patients. We shall consider the reasons for this belief later on, in the discussion section. Our belief that there would be loss in predictive efficiency led us to try out their formula on a new sample of patients.

Procedure

Subjects. The subjects are 33 patients who received individual psychotherapy in the Psychiatric Outpatient Clinic at the New Haven Hospital. All patients who received individual psychotherapy during the period from July 1949 through June 1952, who were treated by a regular staff member (psychiatric resident) of the clinic or by a psychology graduate student under close supervision, and who took the Rorschach test at or near the beginning of treatment (before the ninth treatment interview) are included in our sample. Two-thirds of these patients were diagnosed as neurotic; the remaining third were diagnosed as suffering from psychosis or character and behavior disorder.

Not all patients seen in the clinic during this period were tested, since the psychologists' work load permitted testing a patient only when the therapist specifically requested it.

Assignment to continuing or discontinuing group. A patient was assigned to the "continuing" group if he had kept at least 9 appointments for psychotherapeutic interviews or if he had quit before the ninth interview with the consent of the therapist. Otherwise the patient was assigned to the "discontinuing" group. There were 21 patients in the continuing group and 12 in the discontinuing group.

In counting the number of interviews, we started with the one immediately following the test administration. Interviews before the testing did not count toward the total of nine. Practically this did not make any difference because no patient who was assigned to the discontinuing group had as many as 9 interviews whether or not the pretesting sessions were counted.

Recording of Rorschach scores. Records were scored according to the procedures described by Klopfer and Kelley [5]. The original scoring was reviewed by the authors who reached a consensus on the scoring of all the responses. The total number of responses, the number of large-usual-detail responses, the number of FC responses, and the number of CF responses were recorded for each patient.

Computation of Y score. A Y score for each patient was computed according to the Kotkov-Meadow formula.

Results

The Kotkov-Meadow formula failed to predict which of the 33 patients continued in therapy and which did not. When patients with a Y score of .01132 or higher were assigned to the "continuing" group and those with lower scores to the "discontinuing" group, 17 out of 33 patients were correctly classified (52%). Obviously, this is no better than we could do by tossing a coin or throwing dice. The biserial correlation between Y score and the criterion was $+.19$, which is not significant.

---

[2] Personal communication from Dr. Meadow.

Although the Kotkov-Meadow formula did not predict continuance of psychotherapy among our group, there was a chance that one or two of the scores utilized in the formula would do so. Accordingly, we tested the correlation of R, $D\%$, and $FC - CF$ with the criterion. The chi-square test showed only R to be significantly associated with continuance in therapy ($\chi^2 = 7.14$, with 1 df; $p < .01$). When patients were assigned to the "continuing" group if they gave 19 or more responses (18 was the median number for all 33 patients), 22 of the 33 patients were correctly classified (67%).

In order to test whether there was a continuous relationship, over the whole range of R scores, between R and perseverance in therapy, we applied Festinger's d test for rank-ordered data [2]. This test does not involve any assumption about the form of the distribution function. According to this test the association between number of responses and continuance in therapy was not statistically significant ($p = .11$).

The association between $FC - CF$ score and continuance in therapy is in the opposite direction from that predicted by Meadow's formula. In our sample of 33 patients, a low $FC - CF$ score is associated with continuance in therapy. This relationship is not statistically significant ($\chi^2 = 3.79$, with 1 df; $p < .10$).

The use of IQ score as a predictor. Since various studies [10, 13, 14] have shown a correlation between number of Rorschach responses and intelligence, we believed it possible that the relationship between number of responses and continuance in therapy only demonstrated that more intelligent patients are more likely to continue. To test this hypothesis, we computed the correlation between number of responses and continuance in therapy, with IQ score partialed out. The tetrachoric correlation between R and continuance was + .55; but when IQ (as measured by verbal scale of the Wechsler-Bellevue) had been partialed out, the correlation dropped to + .07. Apparently the relationship between R and continuance was entirely accounted for by their correlations with verbal IQ. The IQ score gave better predictions of continuance than the number of Rorschach responses. The biserial correlation between IQ and continuance was + .71 (significant at .05 level). Since IQ scores were not available for all the patients, these correlations were all computed from a sample of only 23 patients.

Discussion

Why did the Kotkov-Meadow formula fail to predict continuance in psychotherapy in our sample? Before we can answer this question, it is necessary to consider the reasons why cross-validation and validity-generalization studies are necessary.

The need for cross validation and validity generalization. When a prediction formula is applied to the sample from which it has been derived, it is bound to work well. When applied to a new sample, it often works less well. There are four reasons for this:

1. A multiple correlation, like all correlation coefficients, is subject to sampling error. The sampling error of the multiple correlation coefficient is, however, greater than the sampling error of a simple correlation coefficient. Thorndike points out, "Even if the correlations with the criterion are all zero in the population, some prediction will be possible within the sample. . . . The more predictor variables, the higher the multiple correlation which one may expect to derive from them, in the absence of any genuine relationship" [12, p. 203]. Since Kotkov and Meadow tried out about 30 Rorschach scores before selecting 3 for their formula, their formula in effect is based on 30 predictor variables, 27 of which are given zero weights and 3 of which are given positive weights.

2. Thorndike also reminds us that "in addition to any true relationship that may exist between the test variables and the criterion in the parent population, a prediction based on a few chosen tests capitalizes on chance fluctuations in validities and intercorrelations of the variables in the specific sample. . . . The smaller the number of variables selected to be weighted, the more premium is placed upon favorable chance fluctuations in those particular variables retained and weighted. It must be anticipated, therefore, that the multiple correlation obtained by weighting only a picked few of the tests of a battery will show a marked shrinkage, *greater* than when the regression weights are based on all the tests" [12, pp.

204-205]. Kotkov and Meadow gave greater-than-zero weights to only 3 variables. The shrinkage, therefore, is expected to be greater than it would be if they had retained all 30 variables. Since, however, they probably gave zero weights to the other 27 variables because these variables would have had very small weights, the effect of selecting only 3 out of 30 is probably much less important than the effect of trying out so many variables (the reason for shrinkage discussed in paragraph 1 above).

3. Quite apart from the two statistical reasons just discussed, a multiple correlation coefficient may shrink on cross validation because the new sample does not resemble the old in every relevant characteristic. In this imperfect world, successive samples can seldom be identical in every important way. Therefore it cannot be strictly true that successive samples are "drawn from the same population." Mosier [9] has called attention to this problem. He deals with the purely statistical reasons for shrinkage in multiple correlation under the heading of "cross validation," and with the shrinkage caused by changes in characteristics of the samples under the heading of "validity generalization." He points out that studies on new and different samples are necessary to establish the limits of applicability of prediction formulas.

4. The likelihood that the positive findings of Kotkov and Meadow — or of any investigators — are partly due to chance fluctuations is increased by the fact that published studies are not a random sample of all studies done. Because of a feeling among psychologists that positive results are more worth while than negative findings, published studies contain an unusually high proportion of positive findings. No one knows how many studies had to be done to turn up the positive results that appear in the published studies.

In the field of clinical psychology, cross-validation studies are relatively infrequent. Probably a prime reason for this is the great value placed by the scientific community on *original* contributions. Repetitions of past work are valued less than new work: this discourages cross-validating studies. Furthermore, clinical psychologists are likely to feel that any failure on their part to reproduce findings of a pioneer worker may be attributed to a lack of clinical skill. This climate of opinion deters clinical psychologists who get negative results from publishing their findings.

Examples of shrinkage in multiple correlation. An example may serve to emphasize the importance of shrinkage in multiple correlation. Buhler, Buhler, and Lefever [1] worked out a "Basic Rorschach Score" to measure adjustment and discriminate between persons who are well adjusted and persons who are poorly adjusted. Although it was not derived by the conventional multiple-regression procedures, the Basic Rorschach Score is essentially a multiple-regression formula. The Basic Rorschach Score worked well on the groups from which it had been derived, but showed a marked drop in discriminating ability when tried out on other groups. In the original sample it discriminated normal subjects (mean score = 21) from neurotics (mean score = 8). But when the Buhlers tried the Basic Rorschach Score on a new sample, it failed to discriminate nurses (mean score = 12) or other normal persons (mean score = 14) from hysterics (mean score = 12). The Buhlers then maintained that the nurses and other normal people of the new sample were after all not so normal as the original group of normal subjects. We believe, however, that the failure of the Basic Rorschach Score to discriminate in the new sample illustrates shrinkage in multiple correlation.

Kurtz [8] also reported a dramatic shrinkage of multiple correlation upon cross validation. A list of 32 Rorschach signs correctly classified 79 out of 80 insurance agency managers as successful or unsuccessful — in the sample from which the sign list had been developed. When the signs were applied to a new sample of agency managers, the correlation between sign score and success dropped to + .02.

Why did the Kotkov-Meadow formula fail in the New Haven sample? Since Kotkov and Meadow did attempt a cross validation of their findings — strictly speaking a validity-generalization study — and found that their formula also worked in the new sample, it seems unlikely that the failure of the formula when applied to our sample can be entirely attributed to sampling error. Although this seems unlikely, part of the loss in predictive efficiency may nevertheless be laid to this source of shrinkage. If the fact that the formula worked for them but not for us cannot be wholly explained by sampling error, it can probably be explained by differences between the patients seen in the Boston VA Clinic and in New Haven Hospital. The patients in our study, however, seem to be as much like those seen in the Boston clinic as patients in different clinics usually are. Therefore, we have discovered an important limitation to the generality of Meadow's formula. Our finding that the formula does not work in the New Haven sample thus illustrates the importance of the type of study that Mosier called "validity generalization."

Alternative explanations. It is possible, of course, that a VA facility gets patients who differ from non-VA patients in motivation for psychotherapy. In a private-hospital clinic, the fee screens out those with only very slight motivation for treatment. The VA patient is treated free of charge; thus, in a VA facility, some patients with marginal motivation begin treatment. This difference in motivation may affect any attempt to predict continuance of patients in therapy. However, the fact that Rogers, Knauss, and Hammond [11], working in a VA clinic, were not able to predict perseverance in therapy from Rorschach scores, casts doubt on this interpretation.

Another possible explanation of our results is that the treatment procedure is different or the skill of the therapists is different. These might have an effect on the applicability of a prediction formula. In both clinics, the treatment is described as psychotherapy, but this does not guarantee similarity of method. In both clinics, psychotherapy was carried out by therapists believed to be competent, but therapists in one clinic may have been much more skilled than those in the other. We cannot decide from the available information whether such differences might account for the differences in findings. We can say, however, that if such differences in method and skill do affect the usefulness of the prediction formula, then the generality and usefulness of the formula is quite limited.

Implications of this study. Our validity-generalization study shows that the Kotkov-Meadow formula is not generally applicable as a method for predicting continuance in psychotherapy. It also calls in question these authors' interpretation of their results. Kotkov and Meadow believed that patients who continued in therapy did so because they had a greater capacity for bearing the anxiety aroused by psychotherapy. They thought that the $FC-CF$ score was a measure of this capacity for bearing anxiety — as Rorschach workers would say, a measure of "intellectual control" of emotions. Our findings that the patients who quit have higher $FC-CF$ scores would seem to show either that $FC-CF$ is not a measure of intellectual control or that intellectual control is not related to continuance in therapy.

Our study does confirm one of Kotkov and Meadow's findings. Like them, we find that the total number of responses (R) is associated with continuance in psychotherapy. The association is not continuous over all values of R; i.e., we do not find that the more responses a patient gives, the better chance there is he will stay in treatment. Rather, we find that he is more likely to stay in treatment if he gives at least a "sufficient" number of responses, i.e., if he gives more than 18, which is the median number in the whole sample. One plausible interpretation of this is that giving a "sufficient" number of responses indicates a cooperative attitude and that the patient who is motivated to cooperate in the testing is also motivated to cooperate in the psychotherapy.

Our finding that the Wechsler-Bellevue verbal score is, in our sample, a better predictor of persistence in therapy than the number of Rorschach responses suggests that this test may be more useful than the Rorschach for predicting continuance in therapy. Such a use of the Wechsler-Bellevue verbal scale should, however, be supported by evidence from cross-validation studies before it is accepted as valid.

Built-in cross validation. The present authors urge the general adoption of an ingenious method of "built-in cross-validation" used with good success in the AAF Psychology Program. To use this method [Guilford, 3, p. 884], the investigator divides his original sample into two random halves. He develops a multiple-regression formula from the data of one-half of the sample and then applies it to the data in the other half. This method insures that error variance will not inflate the coefficient of multiple correlation. It does not, however, take care of the shrinkage of multiple R attributable to differences in the characteristics of successive samples. Only a validity generalization study, making use of an entirely new group, can do this.

Summary

A trial on a new sample of patients of a formula developed by Kotkov and Meadow for predicting whether patients will continue in psychotherapy was made by the present authors. It was found that the formula, which used three common Rorschach scores: R, $D\%$, and $FC - CF$, failed to predict which of a new sample of 33 psychiatric clinic patients would continue in psychotherapy and which would quit. Although one of the scores, total number of responses (R), was found to discriminate the two kinds of patients, it is felt that this is a measure of the patient's motivation to cooperate or perhaps an index to his functioning intelligence. Both of these can be assessed by more appropriate techniques than the Rorschach test. In fact, we found that all of the correlation between R and continuance can be accounted for by their correlations with the Wechsler-Bellevue Verbal IQ; and the latter test may be a much better predictor of persistence in therapy than any Rorschach score or combination of scores.

The findings of this study thus demonstrate the need for cross-validation and validity-generalization studies. Until such cross validation and validity generalization has been done, no confidence can be placed in the general applicability of any method of diagnosis or any formula for predicting behavior.

Received September 2, 1952.

References

1. Buhler, Charlotte, Buhler, K., & Lefever, D.W. *Rorschach standardization studies. No. 1: Development of the Basic Rorschach Score.* Los Angeles: Authors, 1948.
2. Festinger, L. The significance of difference between means without reference to the frequency distribution function. *Psychometrika,* 1946, **11**, 97–105.
3. Guilford, J. P. (Ed.) *Printed classification tests.* A. A. F. Aviation Psychology Program Research Report No. 5. Washington, D.C.: U. S. Government Printing Office, 1947.
4. Kelly, E.L., & Fiske, D.W. *The prediction of performance in clinical psychology.* Ann Arbor: Univer. of Michigan Press, 1951.
5. Klopfer, B., & Kelley, D. McG. *The Rorschach technique.* Yonkers, N.Y.: World Book Company, 1946.
6. Kotkov, B., & Meadow, A. Rorschach criteria for continuing group psychotherapy. *Int. J. Group Psychother.,* 1952, **2**, in press.
7. Kotkov, B., & Meadow, A. Rorschach criteria for predicting continuation in individual psychotherapy. *J. consult. Psychol.,* 1953, **17**, 16–20.
8. Kurtz, A. K. A research test of the Rorschach test. *Personnel Psychol.,* 1948, **1**, 41–51.
9. Mosier, C. I. Problems and designs of cross-validation. *Educ. psychol. Measmt,* 1951, **11**, 5–11.
10. Neff, W.S., & Lidz, T. Rorschach pattern of normal subjects of graded intelligence. *J. proj. Tech.,* 1951, **15**, 45–57.
11. Rogers, L. S., Knauss, Joanne, & Hammond, K. R. Predicting continuation in therapy by means of the Rorschach test. *J. consult. Psychol.,* 1951, **15**, 368–371.
12. Thorndike, R. L. *Personnel selection.* New York: Wiley, 1949.
13. Wishner, J. Rorschach intellectual indicators in neurotics. *Amer. J. Orthopsychiat.,* 1948, **18**, 265–279.
14. Wittenborn, J. R. Certain Rorschach response categories and mental abilities. *J. appl. Psychol.,* 1949, **33**, 330–338.

Journal of Consulting Psychology
Vol. 17, No. 2, 1953

Rorschach Scoring Categories as Diagnostic "Signs"[1]

Martin Berkowitz and Jacob Levine
Veterans Administration Hospital, Newington, Conn.

The use of certain scoring categories in the Rorschach for discriminating between the "normal," neurotic, and the psychotic has considerable appeal, particularly for the clinician who is seeking a dependable screening technique. As a clinical procedure, these Rorschach "signs" would have much to recommend them, if their validity in differential diagnosis were well established. But as yet few attempts have been made to determine on a proper statistical basis just how effectively these "signs" are able to discriminate between the major psychiatric groups.

In their original study, Miale and Harrower-Erickson [7] found that 9 such "signs" could discriminate between the neurotic and the "normal." These were: (*a*) 25 or fewer responses; (*b*) not over 1 *M*; (*c*) more *FM* than *M*; (*d*) color shock; (*e*) shade shock; (*f*) refusal to respond on any card; (*g*) over 50% *F* responses; (*h*) *A*% over 50; (*i*) not over 1 *FC*. These investigators observed that the presence of 5 or more of these "signs" "suggests strongly the presence of psychoneurosis." Harrower [3, 4] later confirmed these findings.

The question arises whether or not these same "signs," now identified as "neurotic signs," are equally able to differentiate between the neurotic and psychotic. The difficulty in differentiating these two groups on a clinical basis is not nearly so great as it is between the neurotic and the "normal." Harrower-Erickson states "... we have found in a preliminary study with 30 psychotic patients that while a greater percentage than in the case of normal controls will be shown to have over 5 signs, the number will not be nearly so high as in a group of neurotic patients. Moreover, when the incidence of the individual signs shown in the psychotic records is plotted against the normal or the neurotic, the curve will be found to have its own pattern in contradistinction to that of the neurotic or normal" [3, p. 113]. In view of the fact that these interesting findings are the result of a preliminary study with no data or statistics reported, it appears desirable to repeat the study.

The present study is therefore an attempt to test the validity of these 9 Rorschach scoring categories as differentiating "signs" between psychoneurosis and psychosis. Further, an attempt will be made to determine whether or not other scoring categories, particularly *F*+% and number of Populars, are more sensitive differentiators than the others. Clinicians generally use the latter two categories as important indicators of psychosis. On this basis, we might expect that *F*+% and number of Populars will discriminate in a statistically significant way between the psychotic and the nonpsychotic.

Two groups of psychiatric patients from two Veterans Administration hospitals comprised the subjects of the present study. One group consisted of 25 neurotic patients who were randomly selected from the open ward of the hospital. The diagnosis of psychoneurosis was made by staff psychiatrists after careful review of the clinical findings and a number of interviews. The second group consisted of 25 psychotic patients who were also randomly selected from the closed ward of a second Veterans Administration hospital. All were clinically diagnosed as schizophrenic.

The Rorschach protocols of both groups of patients were thoroughly reviewed to insure consistency of scoring. Color and shade shock were evaluated according to the criteria of Brosin and Fromm [2]. Klopfer's classification of popular responses was used. Five or less popular responses was taken as a psychotic "sign" on the basis of Klopfer's statement that "the use of 5 or more popular concepts seems to assure that the subject possesses capacity and interest in thinking along the same lines as other people in sufficient degree" [6, p. 216]. The cutting score for $F+\%$ as a psychotic "sign" was taken as 60 which Beck [1] states is the "critical minimum."

---

[1] Submitted with the approval of the Chief Medical Officer, Veterans Administration, who assumes no responsibility for the opinions expressed here, which are those of the authors.

The differences in the number of neurotic and psychotic "signs" which were obtained from the present two samples were treated for significance by means of a test designed to measure the significance of the differences between percentages [5]. The significance of the difference between the neurotic and the psychotic groups with respect to mean absolute frequency scores of these "signs" were determined by the use of the *t* test.

Table 1

Comparison of "Neurotic Signs" between Neurotics and Psychotics in Terms of Percentages

"Signs"	Total Signs		Percentage		
	Neurotic	Psychotic	Neurotic	Psychotic	Significance of Difference
1. 25 responses or less	20	14	80	56	N.S.*
2. 1*M*	13	12	52	48	N.S.
3. *FM* > *M*	11	13	44	52	N.S.
4. Rejection	11	6	44	24	N.S.
5. *F*%>50	15	11	60	44	N.S.
6. *A*%>50	9	10	36	40	N.S.
7. 1 *FC*	17	18	68	72	N.S.
8. *CS*	16	16	64	64	N.S.
9. *SS*	12	19	48	76	.05
5 or more neurotic signs	18	15	72	60	N.S.

* N.S.—not significant.

The results of the present study are summarized in four tables. As can be seen in Table 1, there are no statistically significant differences between the neurotic and psychotic groups with regard to 8 of the 9 "neurotic signs." With respect to the ninth "sign," shade shock, the difference is significant at the .05 level.

Table 2

Comparison of Mean Absolute Frequency Scores of "Neurotic Signs" between Neurotics and Psychotics

"Signs"	Mean		SD			
	Neurotic	Psychotic	Neurotic	Psychotic	*t**	Significance of Difference
1. Number of Responses	21.08	26.00	16.45	15.22	1.07	N.S.†
2. Number of *M*	1.72	3.04	1.48	3.27	1.81	N.S.
3. Number of *FM*	2.24	3.16	2.78	2.65	1.16	N.S.
4. Number of Rejections	.76	.68	1.07	1.46	.22	N.S.
5. *F*%	56.28	48.96	21.55	20.71	1.20	N.S.
6. *A*%	46.32	46.12	15.24	17.34	.40	N.S.
7. Number of *FC*	1.16	.92	1.38	1.09	.67	N.S.
Average Number of Signs	4.96	4.76	1.87	2.27	.33	N.S.

* Degrees of freedom is 48 for all categories.
† N.S.—not significant.

In Table 2 the mean absolute frequency scores also show no significant differences between the two groups. Color shock and shade shock are not included in Table 2 because they are not quantified.

Table 3

Comparison of "Signs" of $F+\%$ of Less Than 60 and Fewer than 5 Populars between Neurotics and Psychotics

"Signs"	Total Signs		Percentages		
	Neurotic	Psychotic	Neurotic	Psychotic	Significance of Difference
F+%60*	3	14	12.5	58.33	.001
5*P*	7	18	28	72	.002

* N = 24. In both groups there was one case with an F% of zero.

Table 3 shows that the neurotic group can be differentiated from the psychotic on the basis of the $F+\%$ less than 60, and fewer than 5 Populars as "signs." The significance of the difference is less than .001 and .002 respectively.

Table 4
Comparison of $F+\%$ and Number of Populars Between Neurotics and Psychotics

	Mean		*SD*			
"Signs"	Neurotic	Psychotic	Neurotic	Psychotic	*t*	Significance of Difference
$F+\%$	79.58	60.04	17.28	25.34	2.96*	.01
No. of *P*	4.88	3.84	1.63	1.41	2.36†	.03

* 46 degrees of freedom.
† 48 degrees of freedom.

When we compare the average $F+\%$ and the average number of Populars of the two groups, we find the significance of the difference between the two groups is less than .01 and .03, respectively. These results substantiate those in Table 3.

The results of the present study do not support the use of the so-called "neurotic signs" to differentiate the neurotic from the psychotic patient. The total number of such "signs" does not differentiate these two groups. We find that five or more of these "neurotic signs" is above average for both neurotic and psychotic groups. And if we plot the incidence of individual "signs" for both groups, we do not find a characteristic pattern for each, as suggested by Harrower-Erickson [3]. For purposes of screening or differential diagnosis the use of this procedure therefore appears to be questionable.

We further find that of the scoring categories tested only three show a statistically significant difference between the neurotic and psychotic groups. The psychotics consistently showed greater shade shock, lower $F+\%$, and fewer Populars than did the neurotics. The results show that the established cutting scores of 60 for $F+\%$ and 5 for Populars are effective. These three categories might prove useful as "signs" for purposes of differential diagnosis.

Summary

Two groups of 25 psychiatric patients each, one neurotic the other psychotic, were compared in the performance on the Rorschach with respect to certain scoring categories which are used as "signs." Only one of the nine categories commonly referred to as "neurotic signs" was found to be statistically different for the two groups. The use of five or more of these "neurotic signs" to differentiate between the neurotic and psychotic was not substantiated. A statistically significant difference between the two groups was found with two other scoring categories, $F+\%$ and number of Populars.

Received September 10, 1952.

References

1. Beck, S. J. *Rorschach's test.* Vol. II. *A variety of personality pictures.* New York: Grune & Stratton, 1946.
2. Brosin, H. W., & Fromm, E. O. Rorschach and color blindness. *Rorschach Res. Exch.*, 1940, **4**, 39–70.
3. Harrower-Erickson, Mollie R. The value of the so-called "neurotic signs." *Rorschach Res. Exch.*, 1942, **6**, 109–114.
4. Harrower-Erickson, Mollie R. Diagnosis of psychogenic factors in disease by means of the Rorschach method. *Psychiat. Quart.*, 1943, **17**, 57–66.
5. Johnson, P. O. *Statistical methods in research.* New York: Prentice Hall, 1949.
6. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N.Y.: World Book Co., 1942.
7. Miale, Florence R., & Harrower-Erickson, Mollie R. Personality structure in the psychoneuroses. *Rorschach Res. Exch.*, 1940, **4**, 71–74.

Journal of Consulting Psychology
Vol. 17, No. 2, 1953

Group Psychotherapy With Acutely Disturbed Psychotic Patients[1,2]

Herman Feifel
and Arnold D. Schwartz[3]
Winter Veterans Administration Hospital

Some of the main currents of present thinking with regard to maladjustment in people stress the idea that emotional disturbances have part of their genesis in interpersonal relations, and, consequently, that interaction among individuals can serve as a basic therapeutic agent. This concept, together with the desire to reach more patients than individual psychotherapy permits, has resulted in an increasing use of the group psychotherapy method of approach. The method has been applied with normals, neurotics, and different types of psychotic patients with varying degrees of success. However, a review of the literature points up the fragmentary knowledge and paucity of information existing in the field of group psychotherapy with acutely disturbed psychotic patients.

This paper is a report of an exploratory study of the effects of an open-end type of group therapy on a ward of disturbed psychotic patients in a Veterans Administration Hospital. Since our orientation considers psychosis to be a form of defense reaction to situations of conflict, our efforts were directed toward accelerating the development of self-esteem and strengthening the ability of the patients to cope with factors of social reality. Available data on a patient population similar to the one with which we worked enabled us to make broad statistical comparisons as to the relative effectiveness of the group psychotherapy procedure.

Method

Subjects. One of the authors was assigned as resident psychiatrist to a locked ward on the Acute Section of Winter Veterans Administration Hospital for a period of six months. During the first three months, he used a program of milieu therapy, i.e., occupational therapy, corrective activities, special service functions including gym, attendance at movies and dances, etc. Some of the patients, in this interim, also received hydrotherapy, electroconvulsive therapy, and individual psychotherapy. During the second three months, the ward program was identical for the patients, with one exception — the addition of group psychotherapy sessions.

There were 21 beds on the ward and these were occupied by 100 patients during the six-month period being considered, 51 during the first three months, and 49 during the second three months. Since there was an overlapping of certain patients from the first to second

Table 1
Distribution of Diagnostic Categories in Experimental and Control Groups

Category	Experimental	Control
Schizophrenic, paranoid	14	14
Schizophrenic reaction, unclassified	8	10
Manic-depressive	3	1
Involutional melancholia	1	1
Alcoholic psychosis	2	2
Organic psychosis	2	1
Depression, unclassified	2	1
Psychotic depression	–	1
Character-disorder	2	3
	34	34

---

[1] A portion of this paper was read at the Amer. Psychol. Ass., Washington, D.C., September, 1952.

[2] From Winter Veterans Administration Hospital, Topeka, Kansas.

[3] Formerly a Fellow of the Menninger School of Psychiatry, Topeka, Kansas. Now at Langley-Porter Clinic, San Francisco, Calif.

Table 2
Means and Standard Deviations for Intelligence, Age, and Education of Experimental and Control Groups

Measure	Experimental group ($N=34$)	Control group ($N=34$)
IQ		
Mean	104.3	102.4
SD	5.2	4.2
Age		
Mean	35.4	34.2
SD	11.8	12.6
Education		
Mean	10.4	10.1
SD	2.6	3.0

Table 3
Marital, Urban-Rural, and Race Status of Experimental and Control Groups

Status	Experimental		Control	
	N	%	*N*	%
Single	12	35	14	41
Married	12	35	12	35
Divorced	10	30	8	24
	34	100	34	100
Urban	16	47	19	56
Rural	18	53	15	44
	34	100	34	100
White	33	97	32	94
Nonwhite	1	3	2	6
	34	100	34	100

three-month period, and because we considered only those patients as part of the experimental group who had attended a minimum of eight group psychotherapy sessions, our experimental group comprised 34 patients which was, fortuitously, the number of control patients available to us from the first group.

Tables 1, 2, and 3 present the relevant information concerning the background of both groups. Table 1 indicates that both groups, composed mainly of World War II male veterans, in various stages of remission[4] were well matched in terms of their diagnostic classification. The paranoid schizophrenic and schizophrenic reaction, unclassified, types of category dominate both the experimental and control groups.

The findings in Table 2 show that the groups are also well matched on the variables of intelligence, age, and education. Both are slightly above average in intelligence, have an approximate mean age of 35 years (the age range of the experimental group extended from 21 through 61 years of age; that of the control group from 22 through 64 years of age), and possess an educational level the equivalent of approximately two completed years of high school.

The data in Table 3 illustrate the general similarity of both groups with regard to marital situation, urban-rural background, and racial status. The experimental group has a few less single individuals and a few more divorced persons, and more people with a rural background than does the control group. Except for one Negro patient in the experimental group and two in the control group, all the patients were white and American born.

Analysis of the occupational background of the groups revealed that, in both, skilled workers predominated with a sprinkling of farmers, professionals, and semiskilled persons.

In both groups, approximately 80 per cent of the patients were admitted from outside the hospital, the remainder coming to the ward from other parts of the hospital.

The duration of present hospitalization for the majority of patients in the experimental and control groups was usually less than two months, although a small minority in both groups had been in longer, and frequently as little as two or three weeks. As has been mentioned, a minority in each group received ECT, hydrotherapy, and individual psychotherapy, in addition to the general milieu therapy program and group psychotherapy activities of the experimental group.

Procedure. The meetings were for one hour and took place twice weekly in an informal setting in the ward's dayroom. In all, 20 sessions were held with the experimental group. Chairs were arranged in a circular pattern, the two therapists sitting next to each other and the patients being free to sit in any of the chairs, on card tables, or on the floor, if they so desired. Attendance at the meetings was explicitly made voluntary. The group itself was of an open-end type, i.e., while old members of the group were leaving for other wards, trial visits home, or discharge from the hospital, new additions to the ward were allowed to participate in the meetings as soon as they felt so inclined. Because of the turnover of patients, our last session had only three of the original participating patients.

---

[4] It should be kept in mind that although the majority of our patients were acute psychotics in the sense that this was their first or second "break," a few were now in the exacerbatory stage of a long illness, e.g., one of our manic-depressive patients had been ill sporadically for twenty years.

The group structure was of a permissive nature, i.e., the patients were free to talk about anything they desired, sit wherever they wished, walk out of the meeting if they felt like it, etc. However, since these were acutely disturbed individuals, one restrictive aspect was introduced. The patients were told that no physical assaultiveness or self-destructive behavior would be permitted during the group sessions.

In the matter of recording the sessions we considered various alternatives: wire or tape recording, a secretary using shorthand, or taking notes ourselves. We decided on the last course since it permitted us to secure verbatim notes, as well as get a record of feeling-tones and gestures which might have been overlooked if only verbal productions had been recorded. It also allowed us to note, more specifically, the therapists' reactions to the patients as well as their behavior toward the therapists.

Therapists. The two therapists consisted of the ward doctor and section psychologist who alternated in the roles of group leader and recorder-observer. The orientation of both was to create a friendly and noncritical atmosphere — to foster the patients' introducing as much of the material as possible and their leading the discussions. The therapists did become active under certain conditions, e.g., to help a patient enter into the discussion and to aid him express himself, to promote group interaction, etc. The approach can best be categorized as the "analytic-investigative" orientation described by Standish and Semrad [5].

Rating of improvement. Judgment on the improvement or nonimprovement of patients was made by the ward doctor, Chief of the Acute Section, and Assistant Chief of the Acute Section. At the time of rating the patients, none of the raters were aware that their judgments would be used as a criterion in evaluating the effectiveness of group psychotherapy activities. Although their judgments were made independently, they were probably influenced by previous mutual discussions concerning the hospital needs of the patients. Nevertheless, the very strong agreement with respect to the patients' status in practically all of the cases, we feel, tends to make the role of contaminatory bias less important than it might otherwise seem.

Results

Group Process. The most striking aspect of the experimental group's development was its maturation from a gathering of individuals, most of whose members talked unintelligibly, to a cohesively interacting group animatedly discussing common problems and drawing upon one another for help.

The first meetings were characterized by a good deal of open hallucinating, unintelligible talk, and expressions of autistic thinking, e.g., one of the more disturbed patients stated his concern over what doctors do to patients as he incoherently discussed experiments dealing with the mixing of animal and human blood. Another patient, who had been homicidally assaultive, mentioned, with a smile, that he had many friends whose heads were going to adorn fence posts and that his mother would not escape his wrath either. Testing the new situation was obvious, e.g., one of the patients stated his feeling that the aides were "no damn good, a bunch of incompetents who only worked here because they couldn't do anything else"; another asked, "If we speak to you, won't we stay here longer?" Participation of most of the members during these early sessions was minimal and talk was directed to the therapist rather than to other group members. Also, a good deal of physical activity was noticeable; some of the patients walked around the room, others made frequent trips to the water fountain, and some walked out of the dayroom altogether.

As the sessions progressed, a few of the patients began griping over personal incidents to the therapist, still in a kind of private conversation setup, "When am I going to get my laundry back?" "Why can't I go to the typing class with the open ward?" etc. Little by little, they became interested in how other group members were reacting to what they said, e.g., one of the patients, telling the therapist that he felt responsible for his father's death, turned to the group asking for their judgment about what had happened. Gradually, more and more group members started voicing their feelings. Soon the participants were discussing common personal problems, e.g., "Will my mind be affected because of this sickness I've had?" "Should I tell my employer that I've been in a mental hospital?" as well as common ward problems, e.g., "What should *we* do with those people who spit into the water fountain?" "How about not having any piano playing when some of *us* are trying to sleep?" etc. Exchange of ideas and feelings became frequent and often heated. No longer was the group composed of "passive spectators," but it had become one in which a definite "we" feeling predominated and group members had developed feelings of responsibility toward each other, e.g., one of the patients who had a large farm offered another member of the group a job on it if he couldn't find something in his own line after he got out of the hospital.

A definite by-product of the meetings was the sizeable number of patients who approached both therapists after the sessions requesting individual interviews concerning material they had previously repressed or couldn't talk about.

The general developmental process of our group appears to be similar in many respects to Frank's [2] findings with chronic schizophrenic patients and to those of Luchins [4] with open ward patients in the NP section of a military hospital.

As to the themes dominating the discussions — the early sessions dealt with various expressions of guilt feelings over attitudes and actions, fancied and real, toward close family members. The middle sessions brought to the fore conflicts associated with sexual adjustment, particularly their own sex role and identity, and the problem of relating themselves to women, as well as concern over inability to control aggressive impulses. The last sessions focused on the factors of personal responsibility for their illness and getting well, the stigma of mental illness, and the problems of adjusting outside the hospital.

Some of the major facets that the group psychotherapy experience seemed to provide and accelerate for most of the group members were:

Catharsis — allowing the expression of anger, resentment, and various negative feelings against the hospital and aides, and finally against the ward doctor and therapists. For example, one of the more disturbed patients assaulted one of the therapists during an early session. The question of what prompted the patient to do this and the problem of punishment were brought up immediately and discussed. The group was informed that the patient would be sent to a more secure locked ward so as to protect others from the patient, and the patient from his own fears. The group took three to four meetings to work through their own anxieties before the matter was dropped. The therapists noted that most of the patients identified themselves both with the therapist who had been assaulted, as they individually expressed how they would have handled the situation if the patient had struck one of them, and with the assaultive patient, as they speculated about what should happen to, and what was happening to the patient who could no longer control his aggressive impulses. It was also surprising to note how much of the group interaction was based on "rubbing" and "sniping" at each other.

Communication — increasing the patients' ability to communicate with others and getting them to the point where they could formulate and define their problems. An instance of this is where one of the patients talking incoherently was soon interrupted by another who said, "How can the doctor understand you when I can't?"

Leadership — despite the moving in and out of members and the general flexibility of the group, there was always a core of stimulating participants who "carried the ball." They took the initiative, many times, in responding to the queries directed to the therapists and in organizing the discussions. It was evident, too, that the assumption of group leadership and participation was independent of the age of the patient and based more on individual personality factors.

Insight — many of the group members came to realize that the conflicts they had were of their own making and not just due to "bad luck," "members of their family," or "society's fault." They saw that they would have to adjust themselves to other people and not the whole world to them, and that neither medicine, the hospital, nor the therapists could help them unless they tried to help themselves. This is aptly illustrated by a conclusion the group came to at the end of a discussion concerning ways of handling the international situation, "How can we talk about changing others when we can't even change ourselves?"

Control — strengthening of ego factors was definitely apparent in the increased ability of the group members to direct and respond to strong verbal criticism without being overwhelmed, in their growing ability to identify with other members, and in improved reality orientation.

It should be remembered that the group process described above was by no means a linear development but rather indicates the general direction of the group. Also, the findings were characteristic of the group as a whole and do not imply that each group member maintained this general progress. As a matter of fact, some members of the group became even more disturbed during the period of group psychotherapy.

Quantitative comparison between experimental and control groups. In terms of broad statistical comparison, Table 4 shows the hospital status, after three months on the ward, of the experimental and control groups. It is evident that more of the patients who had the benefit of group psychotherapy improved and less became more disturbed than the group which had not received psychotherapy. In the experimental group, 24 patients showed improvement as against 17 patients in the control group, and only 4 patients in the experimental group became more disturbed as against 7 patients in the control group. No improvement was indicated in 6 patients in the experimental group and in 10 patients in the control group. A chi square was calculated combining the more disturbed and no improvement patients into one "no improvement" category. Yates's correction for continuity was applied (1 *df*). Table 5 indicates the χ^2 to be 2.22 with a p value of 15 per cent. While this is not very significant, the trend is in favor of the experimental group. In addition, analysis of the ward records of both groups indicated that less medicinal and physical sedation, i.e., barbiturates, hydrotherapy, packs, etc., was necessary for the group psychotherapy patients. A sampling follow-up of members from both groups, six months after the present study, revealed that a similar status to the one indicated in Table 4 existed in the improved-disturbed ratio in the experimental and control groups. Unfortunately, we did not have as a control a group that had received no therapy at all so that the effects of natural remission could be evaluated relative to the effectiveness of our milieu and milieu plus group psychotherapy groups.

Table 4

Hospital Status after Three Months on the Ward of Experimental and Control Groups

Status	Experimental *N*	Experimental %	Control *N*	Control %
More Disturbed*	4	11	7	21
No Improvement	6	18	10	29
Improved†	24	71	17	50
	34	100	34	100

*More Disturbed = transferred to a closed ward with greater security precautions.

†Improved = discharged from the hospital, left the hospital on trial visit, or transferred to an open ward.

Table 5

Chi-Square Analysis of Table 4 Combining More Disturbed and No Improvement Categories

Status	Experimental *fo*	Experimental *fe*	Control *fo*	Control *fe*	*df*	χ^2	*p*
No Improvement	10	(13.5)	17	(13.5)	1	2.22	.15
Improvement	24	(20.5)	17	(20.5)			

Qualitative differences between experimental and control groups. Above and beyond the quantitative findings, the authors were strongly impressed by the qualitative differences between the groups.

Relationships between patients in the experimental group were more friendly and easy. The content of "gripe sessions," held once weekly, changed from individual complaints to questions affecting the group. Instead of the usual unproductive griping, these sessions were utilized by the patients to make their ward situation more tolerable and pleasant. Where, at first, the ward doctor led these meetings, one month after the group psychotherapy meetings had been started, the group elected their own spokesman to conduct the meetings. They appointed a secretary who took notes and informed the group of suggestions brought up during the week, action taken on previous suggestions, new members admitted to the ward, etc. They chose an athletic chairman who took the lead in interward volleyball and intraward bowling, etc. In effect, the ward doctor became relegated to a position of advisor to the group. The patients were more than ordinarily reactive to how others viewed them, e.g., one of them found it very difficult to justify his claim that he was in the hospital for a "sprained sacroiliac" when other patients mentioned that it didn't seem to bother him when he was playing volleyball or bowling. Another patient found it increasingly hard to maintain a supercilious air when several patients brought up the point that it seemed strange to them that with all his education, he was on a locked ward just as the rest of them. Further, it was seen that although a patient could feel free to "act crazy" for the doctors, nurse, or aide, he was less inclined to do so before his fellow patients, with the result that psychotic behavior and garbled language decreased. On one occasion the ward patients took to task one of the patients who was always putting up a "phony front" before ward personnel. As a result of the group pressure, the patient's tall tales and protests disappeared after a few days. It was apparent that, in many instances, the opinions and attitudes of the group were more effective instruments in changing behavior of patients on the ward than were the activities of the ward doctor and ward personnel.

From a previous attitude of unconcern toward the ward, the patients in the experimental group developed a definite interest in their ward surroundings. Problems of ward temperature, ventilation, cleanliness, etc., became matters of import. During a meeting of the group, one of the patients brought up the point that the windows and curtains were filthy and hadn't been cleaned in the past six months, and that it probably would take another six months to get them cleaned through the usual channels of red tape. The subject was temporarily dropped. However, a few days later, the patients took the initiative in asking the nurse and aides if they could clean the windows and curtains. The result was a joint effort in which both ward personnel and patients participated.

The ward doctor found his own morale improving and his interest in the individual patient increasing when the patients began to show some change after group therapy had been in progress for a short while. As delusions, hallucinations, and psychotic talk decreased, the doctor felt more at ease on his ward rounds and in interviews, both in discussing the patient's conflicts and in attempting to formulate for himself the dynamic aspects of the problems they presented. The patients themselves also evidenced greater freedom in bringing up meaningful material. In the general permissive atmosphere generated by the group psychotherapy sessions, both patients and doctor were better able to discuss the anxiety attending the patient's ambivalence concerning his leaving or remaining in the hospital, as well as his concrete plans for posthospital adjustment.

The meetings also had an effect on the relationship between ward personnel and the patients. In one of the "gripe sessions," about two months after the inception of group therapy, the patients asked why they weren't allowed out-of-doors in the yard when the weather was good. An aide replied that he could not watch all of the patients when he was the only one on ward duty and that several patients might try to elope. A patient suggested that the group itself take responsibility for informing the aide of any attempted elopement. A few patients stated that they didn't want such responsibility. Further discussion ensued and as a result of a majority vote, the entire group decided to take responsibility. Future events showed that they held up their end of the bargain, informing the aide on several occasions when one or two of the more disturbed patients attempted to get "lost" from the group when going to or returning from a group activity. The patients were able to form realistic relationships with ward personnel on the basis of the kind of individuals they were rather than on the basis of what they thought the ward personnel represented. For example, they complained about one specific aide being lazy because he did not help them with their food trays as he was supposed to do, an opinion which was shared by most of the staff.

The sessions also had their effect on relations between the ward doctor and ward personnel. Before the group therapy sessions with the patients were begun, ward personnel brought up few problems of their own at their weekly meetings with the doctor. After the group psychotherapy meetings were begun, a noticeable change took place in their interest. Nurses and aides began to ask more questions about the patients and to bring up various problems they had in dealing with them. Sharpening of interest was indicated by a developing concern over the proper attitudes they should adopt toward various patients, particularly with regard to handling their anxieties by means other than the usual ones of sedation.

In summary then, perhaps more important than the quantitative improvement in the group psychotherapy patients, in this instance, was the *level* and *quality* of their improvement and its concomitant healthy effects on the functioning of the ward doctor and ward personnel.

Discussion

A characteristic aspect of the group psychotherapy sessions was the manner in which it elicited material which had not come out previously. We think this was due to the group structure setting which stimulated expression of problems and conflicts to a group of peers also having similar emotional difficulties. Because of this aspect, group psychotherapy can often serve as a springboard and valuable adjunct for individual psychotherapy in its grappling with the more deep-seated emotional problems of the patients.

Another characteristic finding was the "proving ground," so to speak, that the sessions offered to most of the patients before they left the ward or hospital, for testing and convincing themselves that they would be able to handle themselves in a less "protected" environment.

We were definitely impressed at the movement of the group as a whole, despite its open-end structure with the constant moving out of old members and the incoming of new ones. It is quite conceivable that positive results might have been intensified had the group been a closed one; the point remains, nevertheless, that forward changes did occur despite changes in the membership of the group.

Another revealing result relating to the homogeneity of groups was our experience that the age differences, per se, of our patients were of little consequence. If anything, the age differences, at times, tended to bring perspective to many of the issues discussed. We do not imply that the age factor cannot be a significant one, but rather that in a group of veterans who have similar emotional problems and past experiences, and interests in common, it does not appear to be as important as in other patients.

The sessions emphasized that far from being seclusive and withdrawn, necessarily, our disturbed patients showed marked sensitivity and reactiveness to other group members. Marked changes in personality were not too obvious in most of the group, but what was patent was the gradual disappearance of asocial tendencies, increased reality awareness, and a reorientation of attitudes, e.g., "We've got to change, not the outside world." Part of the patient's pathology, as Breckir [1] has correctly stressed, disappears when he becomes an interacting group member.

As to technique — we both had to guard against getting too involved with content material. Much better results were secured when we focused on the feelings being expressed by the patients. In attempting to catalyze the discussions, we tried, as much as possible, to make use of the patients' productions and level of discussion as long as we felt comfortable in doing so. We did not confront the patients with past behavior or case-record material except where they themselves brought them up in discussion or by symptomatic behavior during the meetings.

We thought that our sitting next to one

another enabled the patients to equate us and helped them see that each of us assumed the same role at alternate meetings. By leaving the seating arrangements up to the individual patient, a shy patient could progress at his own speed from the periphery of the group (or from the floor behind a chair) to the center of the group next to the leader. Further, a patient who was angry could express his feeling by walking out of the meeting, or by moving to the periphery — a move which lent itself to discussion.

We felt that having two therapists for the group was a decided advantage. At the beginning, particularly, it provided supports that enabled us to weather unpredictable outburts of aggression. The second therapist functioned as an "observer" as well as "recorder" in the sessions he did not conduct. We found this viewpoint to be very helpful during discussions between ourselves dealing with the sessions in pointing out transference and counter-transference phenomena, and guarding against blind alleys. One thing that was specifically noticeable in the early sessions was the conceptualization of the therapists by the patients as parent surrogates with consequent attempts to play one off against the other, utilizing the wedge that one was the ward doctor and the other a "visiting doctor." When the patient realized that this strategy would not work, it disappeared. We also discovered, as Hayward, Peters, and Taylor [3] have pointed out, that the presence of more than one therapist allowed more targets to the patients for expressing ambivalent feelings, and permitted a greater therapeutic force to be exerted.

One of the major questions that comes to mind as a result of this experience is the role played by the factors of natural remission and the sheer "attention" aspect of group psychotherapy as distinct from the procedure itself. Our study obviously does not permit any definite answers. It is quite conceivable that had our experimental group received the benefit, let us say, of a picnic once a week instead of the group psychotherapy sessions, the results would have been the same or, perhaps, even better. Nevertheless, we think that we can legitimately state that "attention" of the group psychotherapy type appears potentially as therapeutically helpful.

We found it difficult to evaluate the effect of concurrent therapies that specific patients were receiving; also to assess properly the relation between time spent in the sessions and the therapeutic results obtained since the patients in our group ranged in participation from 8 through 20 sessions.

It is quite apparent that our study poses more questions than it answers. Research involving more controlled studies of process, objective evaluation of therapeutic effectiveness, as well as extended time-series measurements is a necessity before we can be certain of the validity of any of the observations already made. Some of the main areas that demand future attention are the comparative efficacy of the open-end and closed type of groups, optional versus set topics, the relative effectiveness of group psychotherapy with different types of patient populations, dynamics of the group psychotherapy process and its relation to what occurs in individual psychotherapy, multi-therapists in group psychotherapy, and as has been mentioned previously, the "attention" aspect of group psychotherapy, and the problem of natural remission. One mode of approach, not employed by us, that could be helpful in shedding light on some of these questions would be the use before and after treatment of various measures of personality assessment, like the Rorschach, *Q* technique, etc.

Summary

1. Twenty group psychotherapy sessions of an open-ended structure were held with acutely disturbed closed-ward patients in an NP veterans hospital. They were diagnosed mainly as paranoid schizophrenics and schizophrenic reaction, unclassified. Two therapists, one of whom was the ward doctor, alternated in the roles of group leader and recorder-observer.

2. Despite the continued entrance of new patients, the group, as a whole, moved from autistic thinking and individual preoccupations to concern with common problems and more realistic social relationships.

3. The main themes dominating the discussions were: (*a*) guilt feelings toward close relatives, (*b*) sex adjustment problems, (*c*) control over aggressive feelings, (*d*) personal responsibility for illness, (*e*) the stigma of mental illness, and (*f*) adjustment outside the

hospital.

4. In group interaction, the age factor, per se, was less important seemingly than the similarity of problems and commonness of background in the patients.

5. The sessions were helpful in eliciting material not easily secured otherwise.

6. The patients appeared to give up part of their pathology in becoming interacting members of the group.

7. The presence of more than one therapist probably facilitated the therapeutic process.

8. The method of group psychotherapy is adaptable to treating acutely disturbed schizophrenics.

9. A broad statistical comparison of the group psychotherapy patients was made with another group of patients, similar in background and previously on the same ward with the same physician and an identical treatment program except for the group psychotherapy. The group psychotherapy patients showed more improvement. However, a more impressive finding was the *quality* of the improvement and its reflection in general ward activities and relations with ward personnel.

10. Areas for further research were indicated as well as the necessity for more accurate and controlled studies to substantiate the validity of the present observations.

Received August 27, 1952.

References

1. Breckir, N. J. Hospital orientation and training program for group psychotherapy of schizophrenic patients. *Psychiat. Quart.*, 1950, **24**, 130-143.
2. Frank, J. D. Group psychotherapy with chronic hospitalized schizophrenics. In E. B. Brody & F. C. Redlich (Eds.), *Psychotherapy with schizophrenics.* New York: International Univer. Press, Inc., 1952. Pp. 216-230.
3. Hayward, M. L., Peters, J. J., & Taylor, J. E. Some values of the use of multiple therapists in the treatment of the psychoses. *Psychiat. Quart.*, 1952, **26**, 244-249.
4. Luchins, A. S. Group structures in group psychotherapy. *J. clin. Psychol.*, 1947, **3**, 269-273.
5. Standish, C. T., & Semrad, E. V. Group psychotherapy with psychotics. *J. psychiat. soc. Wk*, 1951, **20**, 143–150.

Journal of Consulting Psychology
Vol. 17, No. 2, 1953

The Treatment of Delayed Speech by Client-Centered Therapy[1]

Henry J. Dupont
George Peabody College

Theodore Landsman
Vanderbilt University

and Milton Valentine
Stanford University

One of the speech problems dealt with by speech therapists is called "baby talk," "retarded speech," or "delayed speech." Berry and Eisenson [2] and Van Riper [7], believe that when an emotional disturbance coincides with the development of speech, there will be speech retardation. This is apparently due to the fact that speech is an overlaid function, with all of the anatomical parts used in speech having a more basic biological function. These parts, the tongue, palate, lungs, etc. are free for speech only when the organism is operating with a certain minimum of dynamic integration.

If the very young child is neglected or rejected and reacts with highly motivated disorganized behavior, it is easy to see how the development of speech could be influenced adversely. The correction of the emotional difficulty might therefore be considered an essential first step in delayed speech therapy.

Delayed speech offers a unique challenge to the psychotherapist who emphasizes the growth aspect of therapy as the client-centered therapist does. If the client-centered therapy, as advocated by Rogers [4, 5, 6], does result in releasing the potential for growth and improved interpersonal relations, then client-centered therapy should prepare the child for speech growth also.

Recently Backus and Beasley, commenting on eight years of experimenting with programs of therapy, have stated that "the changes in speech behavior which took place depended less upon devices for breathing, blowing, tongue exercises, 'ear training' and the like, and more upon forces operating in the interpersonal relationship between child and therapist. . . . [1, p. 4]. It is even possible, as Werner [8] reports, that as the child's emotional and social adjustment improves in therapy, his speech will improve without any treatment of the speech problem as such. This is a plausible prediction because therapy usually results in an improvement in social and interpersonal relations and speech *is* a social skill very essential to social and interpersonal relations. The social milieu of almost every child will contain adequate pressure and stimulation for correct speech, which the child will be able to respond to as his adjustment improves.

It is our purpose in this paper to present and discuss the case history, diagnosis, and treatment of one case of delayed speech. The appropriateness of client-centered therapy as treatment will be discussed and evaluated.

Case History and Diagnosis

Johnny[2] was referred to one of the writers (a speech therapist) by a welfare agency for speech correction. His social worker reported that Johnny was having considerable difficulty both in school and in his foster home. In fact, he was a problem to his social agency, for his speech made it impossible to find a satisfactory permanent foster home for him.

---

[1] A preliminary version of this paper was presented at the annual convention of the Southern Society of Philosophy and Psychology, Section on Experimental Clinical Psychology, Knoxville, Tennessee, April 11, 1952.

[2] In order to protect the child's identity, a ficticious name is used.

When first seen by the therapist Johnny appeared small for an eight-year-old, he was thin, had delicate features, and a very forward and out-going manner. His attempts to speak resulted mostly in vowel sounds, frequently garbled. His hearing as measured by the Maico Clinical Audiometer was normal and he missed only three words out of thirty on the children's version of the Anderson Speech Sound Discrimination Test. A phonetic analysis was made of Johnny's speech for use in studying progress in his speech correction and for diagnosis. Because of his speech, an adequate measure of his IQ was not obtained, but his school performance and his general behavior did not suggest mental deficiency.

A report submitted by the social worker provided the following information: Johnny was the sixth of seven children. When he was three and while his mother was carrying her seventh child, his father was arrested for larceny and sent to jail for eight years. Shortly after this his mother became completely disorganized and unable to care for the children. It was not until one year later, however, that Johnny and his older brother, Tom, were taken from his mother and placed in a temporary foster home. One year and four months later Johnny and his brother were placed in another home supposedly for permanent placement. The new foster parents were unable to accept Johnny because he didn't "talk right," and he and his brother were returned to the temporary home. Tom was soon placed in another home, but Johnny remained in the temporary home where he was living at the time of his referral. From the above, and from his behavior in the interviews, it was felt that Johnny suffered from feelings of rejection to which he reacted with negative and aggressive behavior. A diagnosis of "delayed speech" was made and it was decided that treatment should consist of client-, or in this case, child-centered play therapy. It was also predicted at this time that removal of the emotional problem would result in a spontaneous improvement in speech.

Description of Treatment

Two therapists took part in Johnny's therapy which consisted of forty-one play therapy interviews extending slightly over one year. The play sessions took place in a standard play therapy room.

Because it was virtually impossible to understand Johnny, emphasis was placed on helping him perceive that he was accepted by the therapist. This is an aspect of therapy given emphasis by Landsman [3]. An effort was made to show him that he was still liked and accepted even if his frequent limit-breaking could not be.

A brief description of the pattern of the first several sessions provides a contrast with the last two sessions: In the first several sessions Johnny's behavior followed a consistent pattern of aggression and destruction. In each session, after a brief period of examining the toys, he would begin to laugh and yell with apparent glee as he systematically put the room into shambles.

The last sessions were spent in talking to the therapist about the things he was doing as he played constructively with blocks, pounded nails in a two-by-four, and painted. In the last session, he built a sailboat while discussing the fact that he wasn't coming to therapy anymore. At one point, he reminded the therapist that he hadn't brought more paint as he had promised he would. When the therapist suggested that maybe Johnny was a little angry at him for not keeping his promise he replied, "No, that's all right." At the end of the session he shook hands and said good-bye with no signs of difficulty.

This nonquantitative observation of the improvement in therapy is supplemented by the more objective phonetic analysis made of Johnny's speech by the speech therapist. The before-analysis was made in one of the early diagnostic interviews. The other was made from a recording of the last play session. It must be emphasized that the phonetic analysis was made by a speech therapist who had no part in the treatment.

The Speech Therapist's Analysis of Speech

The before-analysis was as follows: his speech is garbled and unintelligible; he seldom attempts to speak. At its clearest, his speech shows the following omissions and substitutions in all positions: omissions: *s, z, sh, zh, tch, dzh;* substitutions: *p* for *l*, *w* for *r*, *f* for *th*, *v* for *th; k* and *t* are used indiscriminately as are *g* and *d*. Most often his speech consists solely of vowel sounds.

The after-analysis was as follows: he initiates speech freely and frequently, he is almost always completely intelligible and only the following errors still occur: *s* and *z* are substituted for *sh, zh, tch,* and *dzh*. *P* is substituted for *l* in the medial position, and *w* for *r* in all positions.

Table 1

Substitutions and Omissions Usually Made in Speech Before and After Therapy

Speech errors	Before	After
Omissions	6	0
Substitutions	6*	10†

*Counting indiscriminate use of K and T, G and D.
†Counting all possible combinations of substitutions.

An example of the way Johnny moved from omission-to-substitution-to-correct-sound would be his speech development on the word *show*. At first he said, "*ow* it to me"; then, "*sow* it to me"; and will soon (we predict) be saying "show it to me."

The social worker reported that both his teacher and his foster mother had noted and commented on his improved behavior and speech. Coincidental with the last several sessions of therapy, Johnny was being prepared for permanent placement in another foster home. This step was now believed practical because of his considerable improvement.

Discussion

As Van Riper [7] points out, in the development of speech the first sounds mastered are the vowels, then the labials, then the dentals and gutturals, then the complicated lip and tongue sounds, then the blends. When Johnny was first seen, his speech had stopped developing at the vowel and labial level. This is approximately the three-year-old level indicating that Johnny's speech had developed little in the last five years.

As shown in Table 1, when last seen he had only the *sh, zh, tch, dzh* blends, and *l* and *r* to master. The disappearance of omission and the increase in substitution show, as the speech therapist reported, that he was more motivated to speak as therapy progressed. This change in motivation to speak is considered a basic factor in the development of Johnny's speech. As is generally agreed, speech is learned and it is therefore important that it be attempted and practiced.

Adequate control is a very great problem in studies of this type. Certainly we do not even approach control of the many variables in the present study. If we were certain that Johnny's speech was due to emotional maladjustment, and if we could be confident that no other factors except the therapy relationship were operating to help Johnny change, then we might conclude with some confidence that therapy without speech re-education is adequate treatment for delayed speech. All that can be said, however, is that it certainly appeared as though Johnny's speech difficulty was due to his negativism and aggression and that as far as we knew there were no other changes in Johnny's life except his relationship with the therapist. If Johnny's speech remained extremely retarded for five years only to improve very much during the one year he was visiting a play therapist, we can at least be suspicious of the therapy as a possible contributor to the improvement. We believe that Johnny's speech improved automatically as he became better integrated as an organism and more efficient in interpersonal and social interactions. Whatever the contributing factors, it is an observable fact that Johnny's speech improved without any speech instruction.

It is dangerous to generalize from one case, but the results of this study tend to re-enforce the hypothesis that child-centered therapy without speech instruction is adequate treatment for some cases of delayed speech. Many additional studies, more thorough than this one, will have to be completed before enough data will be available to prove or refute the present hypothesis.

The results of the study seem to be favorable toward the claims of client-centered therapists regarding the growth-releasing aspects of client-centered therapy, but these claims are by no means proved. The results of the study seem to indicate that therapy can take place even when language communication is difficult and verbalized feelings are at a minimum. Indicating to the child that he was accepted, even when not understood, was apparently adequate to carry the therapy. This is a point made by Landsman [3] and Rogers [4] both of whom stress the importance of accepting the clients' attitudes and feelings even though limits may have to be set on his behavior, especially in the case of children.

Summary

In the foregoing report we have attempted to describe and evaluate child-centered therapy as a treatment method for delayed speech where an emotional disturbance is considered a causative factor. From the description and analysis of one case the following conclusions were indicated:

1. The therapist's observation of the child indicates improvement in emotional adjustment and intelligibility of speech.

2. A comparison, by phonetic analysis, of the speech before and after therapy indicates that improvement occurred without speech instruction.

3. The results of the study are in general consistent with the point of view held by Backus and Beasley [1] except that we would emphasize the importance of being child centered rather than speech centered.

4. The results tend to re-enforce the hypothesis that child-centered therapy can be adequate treatment for some types of delayed speech. The potential value of additional research on child-centered therapy as treatment for delayed speech instead of speech-centered therapy is demonstrated by this study.

Received September 20, 1952

References

1. Backus, Ollie, & Beasley, Jane. *Speech therapy with children.* Boston: Houghton Mifflin, 1951.
2. Berry, Mildred F., & Eisenson, J. *The defective in speech.* New York: Appleton-Century, 1942.
3. Landsman, T. Counseling for emotional problems. *Delaware Sch. J.,* 1951, **16**, 6–7.
4. Rogers, C. R. Significant aspects of client-centered therapy. *Amer. Psychologist,* 1946, **1**, 415–422.
5. Rogers, C. R. Divergent trends in methods of improving adjustment. *Harv. educ. Rev.,* 1948, **18**, 209–219.
6. Rogers, C. R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951.
7. Van Riper, C. *Speech correction.* (2nd Ed.) New York: Prentice Hall, 1939.
8. Werner, L. S. Treatment of a child with delayed speech. *J. Speech Disorders,* 1945, **10**, 329–334.

Journal of Consulting Psychology
Vol. 17, No. 2, 1953

Response to the Human Face as a Standard Stimulus

Ernst G. Beier, Carroll E. Izard, Charles D. Smock,
and Rolland R. Tougas
Syracuse University

In his investigation of the underlying assumptions of the TAT and related projective techniques, Lindzey observed, "Although little effort has been made to explore the variations in fantasy productions between many of the important groups of our own society, it is quite widely accepted or expected that these differences exist. Even such an important cleavage as that between male and female has been little explored so far as TAT behavior is concerned, while such variables as socioeconomic status, occupational role, and ethnic group-membership have also been of slight interest to most investigators" [7, p. 17]. His thoughtful analysis emphasizes the importance of research in this area. Investigations attempting to differentiate among the reactions of various age categories and sexual groups to a standard stimulus may indeed serve to help us to obtain more refined data.

Numerous studies, in the past, have been devoted to problems of differential reactions utilizing the human face as a standard stimulus. For example, the importance of the face as the mediator of expression, of interpersonal responsiveness, and of overt contact with the world at large is generally recognized. Further, at the moment of emergence of social consciousness in the infant, it is likely that the face of the mother (or nurse) serves as stimulus. Recent evidence [4] lays great stress on the cathexis or "imprinting" value of such awareness at the emergent moment, and may well underlie the erotic values attached to certain faces or the repulsiveness of others. Buhler [1] relates that the second month of life features social responsiveness as evidenced by an infant's exclusive reaction to the human face. "The child of two to four months only smiles when it sees the face of a person, but not when its attention is directed to colours, shining objects, or a live cat" [1, p. 159]. Machover [8] in establishing the rationale for the interpretation of drawings of a person, sees the face (or head) as the most representative part of a person. "The face may be regarded as the social feature of the drawing, . . . and is the most expressive part of the body" [8, p. 40]. Further, "The head of the adult is the most important organ relating to the emotional security of the child" [8, p. 39].

Recognition of the importance of the human face as stimulus in social functioning, attitudinal reactions, and early sensitization generally, led several investigators into different directions in search of its intrinsic values. Landis [6] and Frois-Wittman [3] have used the human face in the study of emotional qualities, the social psychologists [10] have used it as a stimulus to elicit prejudices, and Szondi [2] bases his projective technique on value judgments about photographs of human faces, to mention only a few. Through the use of this medium, important results were derived but additional basic knowledge would seem desirable.

The present paper is a report on a study designed to investigate two basic problems related to the human face as a stimulus. The first problem is one of cultural preference, and the second that of differential reaction of the two sexes to the stimuli. For purposes of the present study, the term *cultural preference* intimates the general meaning a stimulus has to large numbers of individuals living within the boundaries of a particular social unit. An illustrative application of the term *cultural preference* may be found when thinking of the facial structure of a movie star; one could predict that a majority of people in our culture would register a "like" response for such a picture. Again, it is known that a large number of persons in our culture prefer the photograph of the only young face in the Szondi pictures. Neither of these facts gives us enough information to formalize our knowledge about the individual who has responded in such a way. In the present study, an attempt was made to throw more light on the problem of cultural preferences. Specifically, which cultural preferences are observable when we break down a large sample of photographs into various age and sex groups? What is, by and large, the feeling toward semblances of older people, those of younger people, toward males and toward females?

Group reactions to as highly structured a stimulus as the human face conceivably conceal large and important differential responses between male and female subjects. In the present study, this possibility was investigated by breaking down the total sample into subparts of male and female subjects. Accordingly, differential sex responses to varying age and sex groups could be determined directly.

It might be argued that in such complex stimuli as photographs of human faces, the individual face would be more important than the specific group to which it belonged and that generalizations could not be made from such findings. It is here contended, however, that the present selection of pictures as well as the large number of pictures used are likely to counteract these difficulties.

Procedure

In order to obtain answers to the questions regarding (*a*) the nature of cultural preferences, and (*b*) the quality of differential responses between the sexes, a large number of photographs of various sex and age groups was obtained and presented to male and female subjects for preference ratings. For purposes of analysis of the data, it would then be possible to break down the stimulus photographs into subgroups according to age and sex and divide the subjects into male and female groups.

Subjects. The sample[1] consisted of 45 males and 31 females enrolled in two sections of a general psychology course at Syracuse University. The age range of the subjects was 18 to 28 with a mean age of 20.41 years.

Selection of photographs. From a pool of over 1000 pictures of human faces, gathered from a variety of sources such as journals and yearbooks, a sample of 250 was selected. Four clinical psychologists served as judges. Selections were made on the basis of the following criteria: (*a*) full-face view; (*b*) nonemotional expressions (no laughter, anger, or fear); (*c*) equal numbers of males and females; (*d*) age groups to cover normal life span; and (*e*) pictures rated unequivocally by the four judges were excluded, to obtain maximum differentiation.

The full-face view was chosen in order to equate the photographs according to structure. The lack of a specific emotional expression was chosen as a selection criterion because it was felt that certain emotional qualities in a face might introduce unaccountable variables. The prejudgment of photographs was undertaken to eliminate such pictures as movie stars and beauty queens. It was felt that this type of picture would be chosen unanimously and thus preclude differentiation. The final pictures chosen reflected certain categories, dividing the photographs into younger, peer, and older age groups.

Table 1 presents a distribution of the selected photographs of human faces according to age and sex.

Table 1
Distribution, According to Age and Sex, of Sample of Photographs of Human Faces

	Age Range				
Sex	1–3	4–16	17–30	40+	Total *N*
Male		29	75	19	123
Female		28	72	22	122
Infants	5				5
Totals	5	57	147	41	250

Administration. The photographs were attached to a strip of tape in random order and then projected on a screen by a standard opaque projector. The size of the picture and the fixation point were kept constant. The screen and seating arrangement were so designed as to present equal opportunity for viewing, and minimizing and equalizing possible distortions. Subjects were presented each picture for five seconds (timed), during which interval judgments were to be made and their responses recorded on the answer sheet, as per the following instructions:

---

[1] The writers wish to express their appreciation to Dr. Roland McKee for his cooperation in making his classes available for this study.

> We have here a series of 250 portraits of people's faces. Obviously, when we are confronted with a stranger, looking at him and at his face particularly, we get certain first impressions about the person. In this research, we are seeking for your first impression. We will present you with the portraits of 250 individuals and some of them will impress you as real likeable people whose company you would appreciate. Some will impress you as people you wouldn't care to meet or associate with at all. Remember, we want your first impression and that we recognize that your first impression would be modified if you knew the person. In between the more extreme likeable and not liked faces, you will find many faces that will impress you as either just o.k. or mildly disliked, both pretty indifferent really. On the answer sheet, we want you to mark *A* for the very likeable; *B* for the mildly likeable; *C* for mildly disliked; and *D* for the definitely disliked. You will have 5 seconds to judge each picture. Each picture must be judged. In other words, each picture should be rated either *A*, *B*, *C*, or *D*. Remember: *A*: a very positive yes; *B*: just yes; *C*: just no; *D*: very definitely no. Remember: First impression only. The time is short — only 5 seconds in which to glance at the picture and then to record your answer. Are there any questions as to what is to be done?

All responses were recorded on IBM answer sheets and machine scored.

In analysis of the data, *A* and *B* categories were combined as "likeable," and *C* and *D* categories were combined as "dislikable." A further breakdown of categories would yield an analysis of the intensity of "likes" and "dislikes" but is not considered part of the present study.

Table 2

Distribution of Preference Ratings by Males and Females on Test and Retest

Measure	Male ($N=45$) Like	Male ($N=45$) Dislike	Female ($N=31$) Like	Female ($N=31$) Dislike
Test	6595	4655	4775	2975
Retest	6731	4519	4778	2972

A retest yielded a frequency distribution of preference ratings as shown in Table 2. An unpublished dissertation [5] has indicated the essential stability of preferential ratings for pictures of human faces as stimuli.

Results

The demonstration of cultural preferences, as defined in this study, may be observed from an analysis of the data in Table 3. The data presented are for the total sample of male and female subjects combined, and are representative only of the 17-30 age group.

For the group as a whole, inspection of Table 3 reveals that only two age-sex groups fail to be differentiated from each other. These are (*a*) peer males (17-30) from older males (40+), and (*b*) older males (40+) from older females (40+). In effect, the subjects divided the categories into two groups; one older than themselves and the other younger than themselves. This finding emphasizes that age

Table 3

Chi-Square Values Comparing Ratings of Sex and Age Groups With Each Other ($df=1$)

	Young Male (4–16)	Young Female (4–16)	Peer Male (17–30)	Peer Female (17–30)	Older Male (40+)	Older Female (40+)
Infants (1–3)	31.77*	16.76*	186.29*	126.72*	157.58*	199.75*
Young male (4–16)		10.53*	338.99*	171.39*	207.38*	290.56*
Young female (4–16)			465.77*	269.51*	294.56*	392.68*
Peer male (17–30)				53.09*	.08	7.84*
Peer female (17–30)					24.62*	60.13*
Older male (40+)						3.77

*Significant at the 1% level of confidence.
Cut-off points for chi square: 6.64=.01 level.
Cut-off points for chi square: 10.83=.001 level.

is an important factor in the measurement of preferences relative to the human face.

The peer-age female category, interestingly enough, was placed with the younger group, and the peer-age male category with the older group. Thus, when considering the peer-age category, age and sex both seem important determinants in preferential ratings, whereas in all categories combined, age alone appeared the more important factor.

Table 4

Cultural Preferences for Age and Sex Groups for All Subjects

	Ratings of All Subjects			
	Like	Dislike	χ^2	p
Infants (1–3) (*N*: 5)	334	46	221.62	.001*
Young male (4–16) (*N*: 29)	1645	559	549.67	.001*
Young female (4–16) (*N*: 28)	1677	451	707.27	.001*
Peer male (17–30) (*N*: 75)	2956	2744	8.13	.01 *
Peer female (17–30) (*N*: 72)	3213	2259	166.56	.001*
Older male (40+) (*N*: 19)	743	801	5.78	.02 *
Older female (40+) (*N*: 22)	802	870	15.51	.001*

*Significant at the level of confidence established for this study.

Up to this point, evidence has been presented to indicate that some age-sex groups are singled out by the subjects for preferential responses. In Table 4, it may be seen that the subjects responded to all age-sex groups in a significantly different manner. The reader will note, however, that in two of the seven groups (older male and older female), the "dislike" responses are larger than the "like" responses, while in the other five categories the "like" responses are larger. These data could be interpreted as indicative of a cultural preference rating. The subjects professed "liking" for photographs of human faces of individuals of peer age and younger than themselves, while they professed a "disliking" for faces of individuals in the category older than themselves. Additionally, large differences were also noted within the five "like" categories. This suggested the possibility of ranking the relevant age-sex categories in order of the observed preference. This study, however, was not designed for such close differentiation, and precise ranking could not be established. The approximate ranking given below might be suggestive of further hypotheses. The ranking was obtained by the following formula:

$$\frac{\text{(Like responses)} - \text{(dislike responses)}}{\text{Number of pictures in each category}} \times 100$$

In this manner, a "like" per picture score was obtained, as shown in Table 5. The ranking would indicate the approximate position of each category in terms of cultural preference.

Table 5

Preference Scores for Age Groups

Age group	Preference score
Infants (1–3)	57.6
Young female (4–16)	21.5
Young male (4–16)	19.0
Peer female (17–30)	6.5
Peer male (17–30)	1.5
Older female (40+)	−3.10
Older male (40+)	−3.05

The contrast in "liking" expressed for children as opposed to the "disliking" directed towards adults is indeed striking. Further investigation of this phenomenon would seem indicated.

Inspection of Table 6 tells us that in four of the seven categories differential responses were found. In their response to infants, to young males, to older males, and to older females, the male subjects respond with a significantly different Like:Dislike ratio than do the female subjects. In their responses to peer-aged men and women, however, as well as to younger girls, the male and female subjects do *not* differ from each other. For all categories combined, men react differently from women in a statistically significant manner. These findings lend support to our hy-

Table 6

Variation of Preferences for Age and Sex Categories of Male and Female Subjects

	Males		Females			
	Like	Dislike	Like	Dislike	χ^2	p
Infants (1–3) (N: 5)	189	36	145	10	7.86	.01 *
Young male (4–16) (N: 29)	930	375	715	184	19.22	.001*
Young female (4–16) (N: 28)	982	278	695	173	1.40	.25
Peer male (17–30) (N: 75)	1741	1634	1215	1110	.25	.60
Peer female (17–30) (N: 72)	1894	1346	1319	913	.22	.60
Older male (40+) (N: 19)	420	435	323	366	4.56	.05 *
Older female (40+) (N: 22)	439	551	363	319	12.77	.001*
Total N:					17.35	.001*

*Significant at the level of confidence established for this study.

pothesis that caution should be exercised in attempting to establish common norms for both men and women in reference to human faces and possibly to all standard projective stimuli.

Discussion

These preliminary findings are reported here because they seem to be of sufficient importance to throw doubt on some of the basic assumptions underlying some present-day uses of the medium under consideration. Definite cultural preferences were shown to be operating in the qualitative responses to age and sex groups through the utilization of human faces as stimuli. The subjects showed more "liking" for faces of peer-age females and younger people, and less "liking" for faces of older people and peer-age males. It becomes clear that more such information will have to be made available in order to fully understand an individual's responses.

These findings seem to mirror a cultural pattern when we discover that older people are the only categories which received a majority of "dislike" ratings. It reflects to some extent the role of the older individual in our society and the sympathy he might expect from a younger group. It is, however, altogether likely that other cultures would demonstrate somewhat different patterns of preferences. Further, it is interesting to note that the male subjects responded differently from the female subjects to almost all age-sex categories with the exception of that representing their own age. Is this an artifact, or have we here an implication of some specific attributes in the development of the cultural-sexual role of an individual?

It was also discovered that responses to human faces by and large differentiate men and women, which indicates that separate norms for the sexes should be considered in order to understand the individual's responses more thoroughly. What remains hidden within normative data which ostensibly differentiate one sex from the other can only be hinted at here. It would seem possible, however, that the social and sexual roles one learns to adopt, the sources and direction of identifications, and the elements of one's self concept may be reflected in qualitative responses to human faces. Machover's monograph [8] alludes to these potentials from an interpretive point of view. That certain faces give expression to one's canalized aspirations or expectancies and thus cue off subjective wishes or identifications would seem pertinent to such age-sex differentiations as were found in this study. This would seem to embody the transferring from previous experience a host of stereotypes and long-standing conditioned responses to such features as the expression, age, and sex of a face. For example, young females have only to be pretty and decorative to command social attention. Again, the need to please others is frequently observed in them. Boys, on the other hand, are expected to make rapid strides in development of physical and sexual power. The influence of such cultural trends on behavior and personality development needs clarification. An unpublished dissertation by one of the authors is attempting to open for consideration the meaning and associational content behind ratings of like and dislike for human faces as a further step in this direction.

Summary and Conclusions

An experiment was designed to investigate the presence of cultural preferences and sex differences in the response of young adults to the human face as a standard stimulus. Subjects for the study were 76 Syracuse University students, 45 male and 31 female. They ranged in age from 18 to 28, the mean age being 20.41 years.

In order to investigate the presence of cultural preferences to human faces, a series of 250 pictures covering infancy to old age in both sexes was shown. The responses of the total sample were statistically evaluated through the chi-square technique. For determination of sex differences, the responses of males and females to each age-sex group were statistically analyzed by the same technique.

The study revealed the presence of clear cultural preferences for faces of people of peer age and younger. These were rated significantly more "liked" than "disliked."

The utility of normative data was shown questionable through the demonstration, within the limits of the particular sample employed, of differential responses by males and females of the same age to pictures of different age-sex groups.

The limitations that these results impose on normative data in response to standard stimuli and possibly all projective stimuli were briefly discussed.

Received August 4, 1952.

References

1. Buhler, K. *The mental development of the child.* New York: Harcourt, Brace, 1930.
2. Deri, Susan. *Introduction to the Szondi Test; theory and practice.* New York: Grune & Stratton, 1949.
3. Frois-Wittman, J. The judgment of facial expression. *J. exp. Psychol.,* 1930, **13**, 113-151.
4. Hutchinson, G. D. Marginalia. *Amer. Scientist,* 1952, **40**, 146-153.
5. Izard, C. E. Perceptual responses of paranoid schizophrenics and normal subjects to photographs of human faces. Unpublished doctor's dissertation, Syracuse Univer., 1952.
6. Landis, C. Studies of emotional reactions, II. General behavior and facial expression. *J. comp. Psychol.,* 1924, **4**, 447-507.
7. Lindzey, G. Thematic Apperception Test: Interpretive assumptions and related empirical evidence. *Psychol. Bull.,* 1952, **49**, 1-25.
8. Machover, Karen. *Personality projection in the drawing of the human figure.* Springfield, Ill.: Charles C Thomas, 1950.
9. Murray, H. A. The effect of fear upon estimates of the maliciousness of other personalities. *J. soc. Psychol.,* 1933, **4**, 310-329.
10. Sherif, M. An experimental study of stereotypes. *J. abnorm. soc. Psychol.,* 1935, **29**, 371–375.

Journal of Consulting Psychology
Vol. 17, No. 2, 1953

Communication and Rapport in Clinical Testing[1]

David Cole
Occidental College

Developments during the past few years in the fields of counseling and psychotherapy have added immensely to our understanding of the nature of communication in the therapeutic situation, and of factors which encourage or inhibit communication between patient and therapist. We have come to regard good communication as prerequisite to good psychotherapy.

The purpose of this paper is to suggest that the insights which we have gained here also have implications in the procedures employed in diagnostic psychological testing. In this area we find that traditionally the emphasis as far as the interpersonal relationships are concerned has been upon rapport rather than upon communication as such. In evaluating the validity of results of such testing we have always been concerned with whether good rapport existed between subject and examiner or whether the performance of the subject seemed to be blocked by the nature of the immediate interpersonal relationship. We have assumed that good rapport was necessary for valid testing.

Most psychometrists will probably accept the hypothesis that a close relationship exists during testing between communication and rapport between subject and examiner. As the level of communication improves, rapport will improve, and conversely, those factors which engender rapport will improve communication.

We are concerned with communication in all types of clinical testing. In the use of projective techniques it is of great importance that the subject feel free to convey to us the responses which occur to him without conscious screening or censoring, and on the other hand that we get a clear picture of what the particular response is that the subject is attempting to describe. Probably many have had the experience in administering the Rorschach of being confronted with a response wherein it was very difficult to ascertain from our subject just how he was using the blot material. We were left with the unpleasant speculation as to whether the apparent fault in perception was his or our own. Again in intelligence testing it is of utmost importance that if we are genuinely to get the qualitatively best performance from our subject, that a high level of communication exist. In general, persons doing intelligence testing for all age levels have followed along the lines of rapport building suggested by Terman and Merrill in the manual for the Stanford-Binet: that "nothing contributes more to satisfactory rapport than keeping the child encouraged," [1], and further that "in general it is wise to praise frequently and generously, but if this is done in too lavish and stilted a fashion it is likely to defeat its purpose." Taking these statements as a starting point, and despite the qualifications injected by Terman and Merrill, persons doing intelligence testing have been prone to use praise as their chief medium of verbal exchange with their subjects, praising profusely and sometimes with too little relation to the qualitative performance of the subject. When a subject failed an item, they followed Terman's suggestion and made "some excuse for it."

---

[1] Paper presented at the West. Psychol. Ass., Fresno, Calif., April 25, 1952.

It is particularly to this approach to rapport building in intelligence testing that this paper is addressed. Our research in psychotherapy has suggested a few things about the value of reassurance in facilitating communication. We have found that in many cases, reassurance, or an evaluative response on the part of the therapist, actually serves as a barrier to further communication. The fact that the response has been evaluated may indicate that all responses will be so evaluated, and the client is thrown on the defensive, setting up a process of screening his statements more carefully before presenting them to his critic. This same process in all probability works during the administration of an intelligence test. Here, with the purpose more often openly evaluative, we consistently find that many subjects enter the testing environment on the defensive. Only the ament or the emotionally infantile is so lacking in autocriticism that he is unable to judge qualitative differences in his responses, and is unable to discriminate success from failure in some test items. Thus, hasty reassurance may often go in conflict with the recognized quality of the responses as seen by the subject, and increase the incidence of blocking and refusal to respond.

The same processes which reduce blocking in psychotherapy are appropriate and desirable in diagnostic testing. If instead of giving fluent and often unwarranted reassurance concerning our subject's performance, we make a sincere attempt to understand what the situation means to him, and attempt to communicate to him this awareness, we may raise the level of communication, and hence get a more valid picture of what kind of a person we are working with. Our aim in testing as far as the establishment of rapport is concerned should be to capture the frame of reference of our subject and to communicate to him our appreciation of his situation.

This approach may influence our administration of standard intelligence tests in a number of ways. With some subjects, the initial reaction to the test will be to be apprehensive, cautious, and defensive. They will indicate this to you through their remarks and behavior as they prepare for the test. Our reaction may recognize this apprehensiveness, letting a subject know that we appreciate how he feels, and can understand why he might feel that way. His comments may be intrapunitive, indicating a doubt in his own ability, "I always do lousy on these things," or may be extrapunitive, "Are these tests really what they are cracked up to be?" He may communicate through behavior more than through words, by sitting rigidly in his chair, arms tightly folded, or by conspicuously and suspiciously eyeing the stop watch. Reacting to such feelings as were just cited with comments such as "you are afraid you never do yourself justice on a test," "you think test scores aren't the whole story," "you seem a little tense," or "you think the timing may bother you," can do much to clear the air and reduce the hierarchical structure which too often exists during testing.

In proceeding through the test, the subject usually first encounters easy items. Various reactions may result. One will find this situation relieving; he is glad that he is able to handle the initial material. This reaction may be recognized; we may indicate that we understand that he is happy to find easy going. Another subject will react to easy items with hostility. He is insulted that we should feel such material was worthy of his abilities. This too can be recognized and accepted as a reasonable reaction from the person concerned. As the testing proceeds most subjects will become aware that the items are more difficult, and that they are finding it more of a problem to produce correct responses. They will react to this in accordance with their own personality and the nature of existing rapport, blaming you, blaming the test, blaming themselves, or accepting the fact that the test is now beginning to separate the men from the boys. In the interest of good rapport and communication it is good to accept these attitudes as expressed, to let the person know that you realize that the items are difficult, that he is uncomfortable (if this is his feeling), and that he will be glad to get out of hot water.

Because most intelligence testing makes an attempt to reach the point where the subject can no longer answer correctly, the wide use of praise can easily lead to a situation where the examiner is confronted with the alternatives of either abandoning his previous approach with the obvious implication that praise is no longer deserved, or continuing to praise in situations where the subject himself is aware that performance is inadequate. To attempt to draw a fine line between praising the subject's effort, and praising the results of his effort, is probably often to exceed the discriminative capacities of both subject and examiner.

The hypotheses and suggestions here presented have proved useful in teaching rapport-building techniques to graduate students in testing courses during the past two years. This approach has been more productive of good rapport than resulted from other methods. As empirical evidence of the nature of the rapport, students using these techniques have been conspicuously more successful in obtaining subjects who are willing to be retested on the alternate forms of the Binet and Wechsler. Inherent in this approach is respect for the person being tested and a recognition of the fact that his feelings are genuine and enter into the testing situation. If what has been found valid in one area of interpersonal relations has validity in another area, then such an attitude may be expected to facilitate communication, rapport, and thus aid in obtaining a more valid picture of the personality under investigation.

Received August 19, 1952.

Reference

1. Terman, L. M., & Merrill, Maud A. *Measuring intelligence.* Boston: Houghton Mifflin, 1937.

Journal of Consulting Psychology
Vol. 17, No. 2, 1953

A Validation of Changes in Scores on the Index of Adjustment and Values as Measures of Changes in Emotionality[1]

Robert E. Bills
University of Kentucky

In a study of the validity of the Index of Adjustment and Values [1], Roberts [2] compared the emotionality of the Index traits which were given high ratings by subjects with traits which were given low ratings. Roberts used the Index traits in free association and his criterion of emotionality was response time. He concluded that the ratings the subjects gave themselves on the Index were valid indices of the emotionality of the traits for the subjects.

The present study was designed to verify Roberts' conclusions and to investigate changes in emotionality when changes in ratings occur from test to retest.

Design

Fifty volunteers, students at the University of Kentucky, were administered the Index of Adjustment and Values at the beginning of a semester. During the following week these students were given, individually, the 49 traits of the Index in free association. The subjects were instructed to respond as rapidly as possible with the first word which "came to mind," and reaction times were measured by means of a chronoscope and voice key. The order of presentation of the stimulus words was randomized and a different order was used for each subject and each of the two presentations. These data were used to verify Roberts' conclusions.

Fourteen weeks after the subjects had been tested with the Index they were retested with this instrument, and during the following week they were retested with the free association test.

Results

Data from the first test. The data from the first Index and the first free association test were used to verify Roberts' conclusions.

The Index of Adjustment and Values requires that a subject make three ratings on a five-point scale for each of 49 traits. These ratings are arranged in three columns which have been designated by the Index authors as concept of self, acceptance of self, and the concept of the ideal self. A fourth score, called discrepancy, is obtained by totaling the differences between concept of self and concept of the ideal self. The three analyses performed in this section related to the ratings on three of these scales.

In the first analysis average reaction times were determined for each subject on the traits he had rated 4 or 5 and 1 or 2 on the concept of self. In this manner two distributions of 50 averages were obtained. Each distribution contained an average for each subject. In all calculations, words on which blocks occurred or on which a subject had a reaction time in excess of 6.81 seconds (3 times the standard deviation of the distribution of mean reaction times of the subjects on each trait) were excluded from the analysis. A test of the significance of the difference between the means of the above two distributions gave a t of 1.111. This may be interpreted, at the .30 level, to mean that there is insufficient reason to assume that the two distributions are not random samples from the same population.

---

[1] The writer is indebted to the University of Kentucky Research Fund Committee. A grant-in-aid from this fund has made the study possible. The writer would also like to thank Glen E. Roberts for his invaluable aid as assistant on this project.

The data were next analyzed for differences in acceptance of self. Average reaction times were calculated for all of the traits rated 1 or 2 (rejection of self) and 4 or 5 (acceptance of self). Again, this yielded two distributions of 50 averages, and each distribution contained an average for each subject. A comparison of the mean difference gave a *t* of 1.88 which is significant at the .06 level of confidence.

The third analysis was for discrepancy scores; the differences between the concept of self and the concept of the ideal self. Average reaction times were determined for each subject on all traits which showed no discrepancy and these averages were compared with the average reaction times of traits which showed a discrepancy. This comparison gave a *t* of 1.80 which is significant at the .08 level. Table 1 summarizes the data from the three analyses.

Table 1
Comparison of Free-Association Reaction Times for High and Low Ratings on Concept of Self, Acceptance of Self, and Discrepancy

Rating	Mean reaction time	Est. SE of the difference between the means	*t*	*p*
Concept of self				
1, 2	2.60	.09	1.11	.30
4, 5	2.70			
Acceptance of self				
1, 2	2.80	.08	1.88	.06
4, 5	2.65			
Discrepancy				
0	2.62	.05	1.80	.08
1, 2, 3, 4	2.71			

Test-retest data. Test-retest data were analyzed to determine if significant differences in reaction times occurred when trait ratings were changed from test to retest. Three groups of analyses were performed on the ratings of concept of self, acceptance of self, and discrepancy.

Four distributions of average reaction times were calculated for each subject on the basis of test-retest ratings on concept of self. Average reaction times were determined for the first and second tests on traits which were rated higher on the retest than on the first test and the same averages were obtained for traits which were rated lower on the retest than on the first test. Thus, four distributions of average reaction times were obtained, and each distribution contained an average for each subject. Similar distributions of averages were calculated for test-retest ratings on acceptance of self and discrepancy scores, and, thus, a total of twelve distributions was obtained. The means of these distributions are presented in Table 2. Subsequently, Index ratings which were higher on the retest than on the first test will be designated "+," and those which decreased on retest "—."

Each of the pairs of "+" and "—" distributions were used to give a distribution of differences. The means of these six distributions of differences are also contained in Table 2. These mean differences represent the change in reaction times for "+" and "—" traits from test to retest and they show that average reaction times were faster on the retest than on the first test.

Table 2
Means and Mean Differences of Reaction Times to Trait Words When Reaction Times Are Arranged According to Increase or Decrease in Ratings on Traits from Test to Retest

Rating	Mean of means 1st. test	 2nd. test	Mean of differences
Concept of self			
Increase ("+")*	2.61	2.51	–.10
Decrease ("—")†	2.70	2.42	–.22
Acceptance of self			
Increase ("+")*	2.67	2.53	–.14
Decrease ("—")†	2.66	2.39	–.27
Discrepancy			
Increase ("+")*	2.77	2.53	–.24
Decrease ("—")†	2.61	2.50	–.11

* Mean reaction times of words which received a higher rating on the Index on retest than on first test.

† Mean reaction times of words which received a lower rating on the Index on retest than on first test.

The above six distributions of differences, three "+" and three "—," were used to calculate distributions of gain scores. Gain scores represent the differences in the changes in reaction times for "+" and "—" traits from test to retest. If the means of these gain scores are significantly different from zero, it may be concluded that changes in reaction time to "+" and "−" traits are significantly different. The means, estimated standard errors, *t*'s, and probabilities for the gain scores are presented in Table 3.

Table 3
Significance of Means of Gain Scores for the Three Index Ratings

Rating	Mean of gain scores	SE of mean	*t*	*p*
Concept of self	−.12	.10	1.20	.30
Acceptance of self	−.14*	.07	2.00	.05
Discrepancy	.13	.06	2.17	.05

*Because of "rounding errors," this mean does not agree exactly with the difference between the mean differences shown in lines 3 and 4 of Table 2.

Discussion

The data confirm Roberts' conclusion that the Index of Adjustment and Values is a valid measure of emotionality, at least as far as can be established by using reaction time in free association as a criterion measure.

The data show, likewise, that changes in ratings of concept of self from test to retest are not accompanied by changes in the emotionality of the traits for the subjects. Roberts [2] demonstrated that ratings of concept of self do not involve distinguishable differences in emotionality and his conclusion is supported by the data of this experiment. This is a result which might be predicted logically since differences in concept of self should not involve emotionality unless a discrepancy exists between the concept of self and the concept of the ideal self or the subject is rejecting of self because of his concept of self. Since different ratings on concept of self do not involve different degrees of emotionality, it was predicted that changes in ratings on concept of self from test to retest would not involve changes in emotionality. The results of the experiment upheld this prediction.

Roberts demonstrated also that comparisons of high and low ratings on acceptance of self, and absence or presence of discrepancy between the concept of self and the concept of the ideal self, revealed significantly different degrees of emotionality. This conclusion was substantiated by the present experiment. Since the variables of concept of self and discrepancy are indicators of emotionality, it was predicted that changes in ratings on acceptance of self or changes in discrepancy scores from test to retest should be accompanied by changes in emotionality. The data of this study support this hypothesis, and thus it may be concluded that changes in ratings of acceptance of self or discrepancy on The Index of Adjustment and Values are accompanied by changes in the emotionality of the traits for the subjects.

The results of the study clearly demonstrate that when ratings are lowered from test to retest, they are accompanied by decreases in free-association reaction time which are significantly greater than the decreases associated with ratings which are raised from test to retest. This is the opposite of what was predicted. It may be reasoned, though, that the ability to lower a trait rating on retest may be indicative of a lower degree of defensiveness, and it would be predicted that such a reaction would be accompanied by a decrease in emotionality and, therefore, by a decrease in reaction time to the trait in free association.

Regardless of the reason, the data permit the conclusion that changes in trait ratings on the Index of Adjustment and Values from test to retest are accompanied by changes in the emotionality of the traits for the subjects. The lack of complete reliability of personality tests has usually been ascribed to sampling errors, errors of measurement, and *perhaps* to changes in the subject. This study has demonstrated that a statistically significant portion of the lack of complete reliability of the Index of Adjustment and Values in a test-retest situation is due to changes within the subjects. These data offer further support for the assumption that the Index is a valid measure of emotionality and changes in emotionality.

Summary

At the beginning of a semester fifty volunteer students were tested with the Index of Adjustment and Values and a free association test which used the traits of the Index as stimulus words. Fourteen weeks later the subjects were retested with both measures.

On the basis of the first tests it was determined that Roberts' conclusions regarding the Index of Adjustment and Values as a valid measure of emotionality were supported.

From the test-retest data it was concluded that changes in trait ratings from test to retest are accompanied by changes in the emotionality of the traits for the subjects and that ratings on the Index of Adjustment and Values are valid measures of changes in emotionality.

Received July 8, 1952.

References

1. Bills, R. E., Vance, E. L., & McLean, O. S. An Index of Adjustment and Values. *J. consult. Psychol.*, 1951, **15**, 257–2[illegible].
2. Roberts, G. E. A study of the validity of the Index of Adjustment and Values. *J. consult. Psychol.*, 1952, **16**, 302–304.

Journal of Consulting Psychology
Vol. 17, No. 2, 1953

The Madeleine Thomas Completion Stories Test

Eugene S. Mills
Whittier College

In 1949 the writer made a survey of fifty psychological clinics and child guidance centers to determine the extent to which the Madeleine Thomas Completion Stories Test was being used in clinical practice. All but two of the clinics reported that they were unfamiliar with the test and about half of them asked for specific information relating to the test and its application. A brief report [1] of this test has been made previously, but it is felt that a further report is warranted by the interested response to the survey questionnaire.

The Madeleine Thomas Completion Stories Test (MTT) was first presented in an article by Madeleine Thomas [3] in 1937. The development of the test was the work of the J. J. Rousseau Institute and of the Psychological Laboratory of Geneva. Helmut Wursten, clinical psychologist for the Children's Hospital in Los Angeles, made a translation of the test which has been used at that hospital and at the Claremont Graduate School, Claremont, California, under Florence Mateer. Using Wursten's translation of the MTT items, the present writer made an exploratory study [2] of the test by administering it to fifty elementary school children between the ages of five and fourteen years. To the best of the writer's knowlege, this is the only study of the test to be reported in a psychological journal since its publication in 1937.

Description of the MTT

The originator of the test offers as a basis for the MTT the hypothesis that all imaginary creation obeys a certain determinism and that one can, when in possession of such a creation, infer the psychological causes from which it springs. The MTT is a clinical tool for releasing this creativity. It consists of fifteen stories or items which relate to the family conditions, school experience, and fantasy life of a fictitious little boy or girl of the same sex and age as the child being tested. Each story poses a problem which is left in suspense, and the child is asked to resolve that problem; that is, he is asked to finish the story to his own liking. The fifteen stories follow:

1. A boy (or girl) goes to school. During recess he does not play with the other children. He stays all by himself in a corner. Why?

2. A boy fights with his brother. Mother comes. What is going to happen?

3. A boy is at the table with his parents. Father suddenly gets angry. Why?

4. One day Mother and Father are a bit angry with each other. They have been arguing. Why?

5a. Sometimes he likes to tell funny stories to his friends. What kind?

5b. Sometimes he likes to tell funny stories to his parents. What kind?

6. A boy has gotten bad grades in school. He returns home. To whom is he going to show his report card? Who is going to scold him most?

7. It is Sunday. This boy has been taken for a ride with Mother and Father. Upon their return home, Mother is sad. Why?

8*a*. This boy has a friend whom he likes very much. One day his friend tells him, "Come with me, I am going to show you something, but it is a secret. Don't tell anybody." What is he going to show him?

8*b*. This boy has a friend whom he likes very much. One day his friend tells him, "Listen, I am going to tell you something, but it is a secret. Don't tell anybody." What is he going to tell him?

9*a*. It is evening. The boy is in bed, the day is ended, the light turned off. What does he do?

9*b*. What is he thinking about?

9*c*. One evening he cries; he is sad. What about?

10. Then he goes to sleep. What does he dream about?

11. He wakes up in the middle of the night. He is very much afraid. What of?

12. He goes back to sleep, and this time has a very nice dream. A good fairy comes to him and says, "I can do anything for you! Tell me what you want. I am going to touch it with my magic rod

and all you may wish for is going to come true!" What does he ask for?

13. The boy is growing up. Is he anxious to be a big boy soon, or would he rather remain a little boy for a while?

14. Among all the fairy tales that have been told him, which one does he like the best of all?

15. Do you remember when you were a little boy? What is the first thing that you can remember now?

Administration and Theme Analysis

The *MTT* is administered in a room which is as free as possible from distracting conditions. After rapport has been established with the child, the examiner begins with words to this effect: "We are going to tell some stories. I will start them and you will finish them. It isn't hard. You will see." Then the examiner reads the first story in a slow, even voice and finishes with an expectant attitude. If the child's response is only perfunctory or seems to leave a great deal unsaid, the examiner uses a follow-up question or comment such as, "Why?" or, "And then what happened?" In this manner, it is often possible to extend the scope of the response to include highly significant material which might otherwise have been left unrevealed. The responses are recorded verbatim.

The MTT requires, on the average, about 20 to 25 minutes for administration. This depends, of course, upon such factors as the speed of response, the length of the response, and the amount of resistance encountered; some children require only about 15 minutes to complete the items.

There are, as yet, no well-established methods for analysis of the MTT; however, methods for processing other types of projective material seem applicable. The writer has used the following general approach with some success. First, a survey of the test responses is made, emphases and trends being noted. Second, each response is studied separately. An attempt is made to determine the dynamics and orientation of the response (i.e., sibling rivalry, directed toward a younger sister; parent-child tension, directed toward the mother). These themes are then recorded in a few words on the test form or on a separate piece of paper along with the story number. Third, the responses are then scrutinized for repetitions, unusual responses, and responses which stand in striking contrast to the general theme developed in the test. Fourth, the most important themes are then studied in relation to any biographical material about the child which is available. In this way, it is often possible to discover many significant facts about the emotional life of the child in a relatively short time, depending, of course, upon the skill and insight of the clinician, and the extent to which he desires to analyze the test responses. In some cases a selective use of certain items has proved helpful, especially when the clinician suspects difficulty in specific areas of the child's emotional life. This flexibility would appear to be one of the major advantages of the technique.

With what ages is the test effective and what kind of material is revealed by its use? The writer has found the test to be useful between the ages of six and thirteen years, with its greatest effectiveness occurring with children from eight to eleven years, inclusive. As with many other tests, much of the effectiveness of the MTT depends upon the responsiveness of the child being tested.

In order of the frequency of occurrence, the writer has found the following themes revealed by the MTT. The numbers following each theme refer to the stories which seem to most effectively elicit that particular theme.

1. Manners and moral conduct—3, 8*a*, 9*a*.
2. Fantasy life—11, 10.
3. Parent-child tensions—3, 6.
4. Parental discipline—2, 6, 9*c*.
5. Anxiety—11, 15, 9*c*.
6. Social adjustment, good and bad—1, 9*c*.
7. Likes—12, 13.
8. Parental conflict—4, 7.
9. Sibling rivalry—2, 3.
10. Aggression—1, 13, 9*c*.
11. Home stability and instability—4, 11.
12. Sex awareness—13, 8*b*, 9*a*.
13. Parent-child understanding—6, 3.
14. Escape—1, 8*b*.
15. Fears—13, 11, 6.
16. School conduct—1, 9*b*, 12.
17. Teacher-child relationships—1, 6.
18. Deep loss—15, 10.

Other significant indications which may appear are resistance and distraction. While not directly revealing the nature of the child's emotional life, these indications are often significant when one is cognizant of the specific story-items to which the child is responding.

It is hoped that this brief description will serve to acquaint clinicians with the Madeleine Thomas Test and perhaps stimulate further research.

Received August 26, 1952.

References

1. Mills, E. S. The Madeleine Thomas Test as an aid in reading children. *Fourteenth Yearbook,* Claremont College Reading Conference, Claremont, California, 1949.
2. Mills, E. S. A study of the Madeleine Thomas Completion Stories Test with fifty elementary school children. Unpublished master's thesis, Claremont College Library, Claremont, California, 1949.
3. Thomas, Madeleine. Methode des histoires à compléter. *Arch. Psychol., Genève,* 1937, **26,** 209–284.

Journal of Consulting Psychology
Vol. 17, No. 2, 1953

MMPI Profiles and Personality Characteristics[1]

H. Birnet Hovey
Veterans Administration Hospital
Fort Douglas Station, Salt Lake City, Utah

This study is concerned with comparing observed personality characteristics of nonclinical individuals with profiles they produced on the Minnesota Multiphasic Personality Inventory (MMPI) [1]. Contrary to the usual procedure of comparing test results with groups of individuals selected according to certain behavior criteria, this study started by grouping persons according to test results, and then ascertaining what personality characteristics or traits appeared, a procedure similar to the one used in compiling *An Atlas for the Clinical Use of the MMPI* [2]. The population consisted of student nurses in practicum training, not because of a primary interest in nurses, but because data for the most part were already available and only waiting to be organized and analyzed. The population consisted of individuals who were making constructive social and vocational adjustments, who had been under discerning supervision of experienced psychiatric nurses and other staff personnel, and for each of whom anecdotal records with expressions of behavioral and personality characteristics had been made independently by eight of these personnel.

The program in the Veterans Administration Hospital at American Lake, Washington, carries on an intensive practicum for student nurses, admitting a class of from 16 to 24 students at a time, for a three-month period. Within the first day or two, the MMPI is administered with other tests and interviews. Then the students are separated into small groups and assigned to four supervisors in rotation. These supervisors record impromptu impressions of each student every few days, in addition to using a rating scale. Also, four additional staff personnel, under whom each student is observed in special assignments, record comments. These various notes are sent to central files.

Procedure

Copious notes and MMPI answer sheets were available for study on 97 former students. The MMPI scores were computed, using *K* (and omitting *Mf*). There turned out to be 92 personality characteristics mentioned by at least two supervisors in describing students. One-half represented assets, while the other half represented liabilities approximately opposite to the assets.[2] Tetrachoric correlation coefficients (*tr*) were computed between scores on MMPI scales and frequencies of characteristics or traits as mentioned by supervisors. For one scale at a time, the distribution of subjects in terms of scores was divided dichotomously at the 75th percentile. Whenever a dividing point could not be made close to this percentile because of a few subjects having the same score in that region, these few cases were separated according to the relationship of subject's score on that scale to her own mean score for all the scales. (For instance, if three subjects had identical scores cutting the distribution to one side or the other of the 75th percentile on a scale under consideration, and one of the cases was needed in the high-scoring group to round it out closer to a quarter of the distribution, the subject with the lowest general profile level would be included as her score on that scale would be relatively higher than those of the other two subjects in terms of their over-all profiles.) In computing an *r*, the top 25%, or high score, was compared with the remainder of the distribution for proportions in which a trait occurred. The same procedure was used for low score, this time using the 25th percentile as the dividing point. Divisions were made near the ends of distributions rather than at the centers, because of interest in what the more extreme scores might represent, analogous to an interest in superior intellect or in mental deficiency.

---

[1] From the Veterans Administration Hospital, American Lake, Washington. Dr. James C. Stauffacher, Chief Clinical Psychologist there, gave valuable suggestions in the preparation of this study, and Prof. Allen L. Edwards, University of Washington, gave essential advice on some of the statistical problems.

[2] Whenever a given asset had been mentioned by one supervisor and its approximate opposite by another supervisor, neither was used. This occurred infrequently. Far more frequently the items had been mentioned by at least two supervisors. In fact over 40% of all the entries used in the study were supported by the same entries on the same individuals being made by at least two different supervisors.

Comparisons were made for high and low scores on all eight of the clinical scales, with each of the 92 traits. High and low *L*, *K*, and mean score were also included. There were too few *?* scores above zero, and the spread in *F* scores was too narrow to include these in the comparisons. For each scale and in both directions, the traits were listed which correlated with it at or above the .05 level of confidence. Coefficients which failed to reach this level were dropped from further consideration. In addition to traits, the MMPI scales were compared with elements evaluated by the supervisors on the formal rating scale.

These procedures were repeated on a new group of 40 student nurses. Only correlations between MMPI scale scores and personality characteristics which reached the .05 level or better both times were considered as sufficiently significant for reporting here. The new group contained too few low *K* scores for use in computations.

Results

For the 92 traits considered in this study, there was a possibility of 1932 coefficients of relationship, of which 115 turned out to have *tr*'s at the .05 level of significance or higher, for both the original group of 97 students and the new group of 40 students. Below are listed deviations in the positive and negative directions for the various MMPI scales, and the related traits. The traits are listed in descending order according to size of *tr*. The chi-square test (χ^2) was applied, and those items which reached the .05 level or better for both the original and new samples are italicized. (Some findings not supported by χ^2 nor listed are discussed in footnote 3.) An asterisk (*) denotes that the relationship is supported by a significant one, at the .05 level, in the opposite direction for the group scoring opposite on the same scale. All remaining items in the list are similarly supported but not significantly so. When a trait in the list is followed by "—minus," a negative *r* is denoted. However, when significantly positive and negative relationships obtained for traits with the two extremes of scales respectively, only the positive ones are listed, the negative ones being implied by the asterisks.

High *L* score: **Afraid of mental patients,* *Has difficulty expressing self orally, *Persevering,* Passive, *Feels insecure about self

Low *L*: **Poised and at ease around others,* Not industrious, *Friendly*

High *K*: Reserved, *Mature,* Afraid of mental patients, Has difficulty expressing self orally, Friendly

Low *K*: (Too few cases in the new group for statistical treatment)

High *Hs*: *Adjusts slowly, *Alertness—minus, Ease of oral expression—minus*

Low *Hs*: (None other than opposite of first item for High *Hs*)

High *D*: **Shy,* *Nonaggressive

Low *D*: **Initiative,* **Poised and at ease around others, Participates actively in group discussions,* *Adjusts rapidly, *Good socializer, **Emotionally stable, Efficient,* *Ease of oral expression, Desires responsibilities

High *Hy*: *Immature,* **Friendly, Enthusiastic,* Careless in personal appearance, *Cooperative, Cheerful

Low *Hy*: Industrious—minus, Self-confident

High *Pd*: **Participates actively in group discussions,* **Initiative,* *Aggressive, *Desires responsibilities,* Not industrious, Adjusts rapidly, Self-confident, Shy—minus, *Unafraid of mental patients, Works persistently with assigned patients—minus, Enthusiastic

Low *Pd*: *Suggestions accepted willingly, **Persevering,* Stimulating personality—minus

High *Pa*: *Lacks self-confidence, *Dependent and submissive,* Outgoing—minus

Low *Pa*: *Adjusts rapidly,* Dependable—minus, *Poised and at ease around others

High *Pt*: **Participates little in group discussions,* Poor socializer, *Shy,* Ingenious—minus, Neat in personal appearance

Low *Pt*: (None other than opposite of first item for High *Pt*)

High *Sc*: *Participates actively in group discussions,*

Ingenious, Initiative, Poor judgment, Not enthusiastic
Low *Sc*: Friendly, Alert
High *Ma*: **Ease of oral expression*, **Initiative*, *Ingenious, *Self-confident, Efficient,* Conscientious—minus, *Reserved—minus,* Responsible, Effective with mental patients, Shy—minus, Immature, Poised and at ease around others, Leadership qualities
Low *Ma*: **Participates little in group discussions,* Adjusts slowly, Persevering
High ave: *Initiative, Ease of oral expression—minus
Low ave: Reserved

In view of the above results, the possibility of predicting final grades achieved in the course from MMPI scores was investigated. High and low deviation on each scale, high and low mean score, wide and narrow spread between scale scores were tried, but significant relationships were not found. The lack of significant relationships might be due to the students having already been screened for motivation and aptitudes connected with nursing and therefore not representing an unselected sample. It is also possible that some of these persons may have adequate counterbalancing assets that enable them to function effectively in spite of certain personality weaknesses, and such assets may or may not be reflected in the MMPI profiles. The lack of success for prediction of grades from profiles was about the same as Weisgerber's [3]. Also, in agreement with his study, there was little success in discovering the students who needed special counseling. A special scale for predicting grades was constructed, based on item analyses of *A* versus *D* students. But when the scale was applied to the new group of 40 students, prediction turned out to be little better than chance.

Discussion

The primary objective of the study was to ascertain what kinds of personality characteristics might be related to high and low scores on the various scales of the MMPI, for a group of normal individuals. The more significant relationships are listed above.[3] Spontaneous notes made by the supervisor-observers were more discriminating than formal ratings made on a rating scale, the spontaneous characterizations correlating more significantly with MMPI scores. In fact, no significant *r*'s emerged for relationships between the scales and the formal ratings. This may be due to the formal ratings' containing more comprehensive judgments, or possibly because they were not on-the-spot observations, or perhaps because they were used for purposes of motivation. Our findings here correspond also with Weisgerber's [3].

The list of characterizations not only reveals differences within one group of individuals but has also shown itself to have predictive value when applied to a new and similar group. So far, the list for predictions has not been used in individual cases, but such a study has been initiated. For groups at least, MMPI profiles can be used to predict personality characteristics which are not necessarily features of emotional illness. However, a word of caution may be in order should one try to predict those characteristics found in our study for other kinds of groups. For instance, in spite of our coefficients reaching significance in both our samples, they might apply only to student nurses who have achieved advanced training status in the Pacific Northwest, and who are observed by the particular supervisors who have been participating in our study.

---

[3] In most instances and as might be expected, a trait correlating with a scale did so in opposite directions for high and low scores for that scale. The list contains these instances. However in some other instances, traits showed up as associated (according to *tr* technique) with deviation scores in one direction only for a given scale. These are not in the list. For instance high *Pa* correlated significantly with "persevering" but there was also a small positive correlation of this trait with low *Pa*. "Friendly" and "ingenious" were associated with low *Hs*, yet high *Hs* was not at all associated with a lack of these traits. High *D* was negatively correlated with "friendly" which means that supervisors noted relatively few of the individuals with elevated scores on this scale as being especially friendly, yet at the same time the high *D* groups did not contain any more than chance expectancy of subjects noted as being on the unfriendly side.

According to *tr* coefficients obtained, a scale elevation might mean either strength or else manifest weakness of the same characteristic. High *Sc* correlated positively with "participates actively in group discussions" and also with "participates *little* in group discussions." When high *Sc* was accompanied by high *Pd* and *Ma*, the individuals concerned tended to participate actively, whereas if neither *Pd* nor *Ma* were high, they were prone not to participate. These relationships held for 13 profiles dominated by the three scales, versus 9 profiles with both *Pd* and *Ma* being several points below *Sc*. The tally was 11 to 1, and 2 to 3, for "actively" to "little," for the two groups respectively. Again, when high *Sc* profiles for 18 individuals who participated actively were compared with 10 who did not, the medians for the latter group were nine points below the former on *Pd* and four points below it on *Ma*. Individuals with elevated *Sc* scores may be considered as having schizoid trends, and they may be inclined either to withdraw from group participation, or else be dominant in the group. One student who obtained a high *Sc* score accompanied by elevated *Pd* and *Ma* scores was characterized as "participating actively," and otherwise was noted to be very aggressively outspoken in class and to change the topic.

The list reveals that some of the traits are correlated with more than one scale deviation. In such instances, the occurrence of two or more of the involved deviations in a sample should increase the proportion of the trait appearing in that sample. There were too few cases of the various kinds of profiles for comprehensive statistical analysis, but one sample tends to bear it out. There were 26 profiles with the two highest scores being on *Pd* and *Ma*, which was the largest number of profiles dominated by any two particular scales. These two scales are correlated with "participates actively in group discussions." The group of individuals producing those 26 profiles contained almost half again as high a proportion of persons with this characteristic as did two groups dominated by *Pd* and *Ma* separately. On the other hand, if a trait is correlated positively with one scale deviation and negatively with another, concurrence of the two deviations should tend to cancel so far as that trait is concerned. "Participates little in group discussions" happens to correlate with high *Pt*. There were 16 profiles dominated by elevations on this scale together with *Pd* or *Ma* but not both. The group producing these profiles contained no more of the trait within it than chance expectancy.

Deviations on every one of the individual scales were more loaded with significant relationships than in mean score or general level of profile. Only three traits correlated with high and low mean score, whereas the various scales excepting *Hs* carried from 5 to 16 trait relationships. Also, none of the items related to high and low mean score is supported by χ^2 while every one of the scales carries one or more such items. This may have something to do with the inability to select effectively from total score students in need of counseling, or to predict grades better than chance. In this study, shape of profile was more fruitful than its general elevation.

Asset traits as well as liabilities may be associated with elevations on MMPI scales, and negatively valued traits may be associated with subaverage scores on the scales. From one to seven positive traits were associated with elevated scores on seven of the ten scales. One positive trait was even associated with a high over-all mean score. Conversely, from one to five negative traits were associated with low scores on all scales but *Hs*, *Pt*, and *Sc*, and also a negative trait was related to low mean score. When only those items supported by χ^2 are considered, the elevated scores are related to twelve "desirable" and eight "undesirable" traits. By the same criterion, the low scale scores are related to thirteen positive traits and six negative ones. Elevated scores in clinical practice tend to signify various kinds of maladjustment or potentials for it, and in view of this, it may be inferred from our results that there are some personality assets associated more with maladjustment potentials as measured by the MMPI than with relative freedom from such potentials.

Summary

A group of 97 student nurses in practicum training had been given the MMPI, and during the training period the supervisors made notes relating to personality characteristics for each student. Tetrachoric *r* and also χ^2 were applied to ascertain any relations between high scores, and also low scores, on the various MMPI scales and observed personality characteristics. The procedure was then applied to a new group of 40 student nurses. Traits are listed which correlated most significantly with one or more of the scales for both groups of students.

Impromptu notes made by supervisors showed more significant relationships with MMPI scale scores than did ratings made on the rating scales. Clinical observations to the effect that for interpretation, deviations on individual scales must be considered along with deviations on other scales, are supported by the data. Individual scores on most scales showed up as more meaningful than general elevation of profile. Some traits of positive value as well as ones of negative value were found to be associated with elevations of various scales, and some negative traits were related to low scale scores. In other words, potential for emotional maladjustment may carry along with it or enhance some positive personality characteristics.

Received September 22, 1952.

References

1. Hathaway, S. R., & McKinley, J. C. *Minnesota Multiphasic Personality Inventory.* (Rev. Ed.) New York: Psychological Corp., 1951.
2. Hathaway, S. R., & Meehl, P. E. *An atlas for the clinical use of the MMPI.* Minneapolis: Univer. of Minnesota Press, 1951.
3. Weisgerber, C. A. The predictive value of the Minnesota Multiphasic Personality Inventory with student nurses. *J. soc. Psychol.*, 1951, 33, 3–11.

Journal of Consulting Psychology
Vol. 17, No. 2, 1953

A Comparison of The WISC and Stanford-Binet IQ's of Normal Children

Glen A. Holland[1]
University of California, Los Angeles

An appreciable literature has begun to accumulate concerning the Wechsler Intelligence Scale for Children. Comparison of this scale with the Stanford-Binet is inevitable. One reason for this is that the Stanford-Binet in its original and revised forms has constituted a standard for intelligence measurement for 36 years. Also, its frequent use in a wide variety of situations has provided individual testers with operational meanings for its scores over and beyond those given by its standardization procedure. An obvious economy of effort is involved if the correlation of the Stanford-Binet with a new scale appears to warrant some degree of generalization from this accumulated experience.

Up to June, 1952, six studies have been reported comparing the Stanford-Binet with the WISC using normal subjects [1, 2, 5, 11], two of these being unpublished Master's theses reported by Pastovic and Guthrie [5] and not seen by the writer. Three studies have been done using defective subjects ranging from the moron through borderline grades of intelligence [4, 8, 9]. The procedures, data, and conclusions reported suggest certain hypotheses. These will be given below in numbered statements along with relevant evidence.

1. The order of testing (S-B first or WISC first) produces a statistically significant difference in the mean IQ's obtained with the WISC. Although this specific hypothesis has not been studied, it is implied in at least four studies in which the investigators report that they have alternated the order of testing [1, 2, 5, 8].

2. The correlations with Stanford-Binet IQ's will be significantly greater for WISC Full Scale and WISC Verbal Scale than for WISC Performance Scale IQ's. Relatively clear-cut results to this effect are given by three studies [1, 8, 11]. Two others have not reported relevant data [4, 9]. Pastovic and Guthrie [5] cite one study in which this hypothesis receives support and three others in which the results are ambiguous. Krugman, Justman, and Wrightstone [2] found that for their total group of 332 subjects the correlations with Stanford-Binet IQ's for the Full Scale and Verbal Scale were higher than for the Performance Scale. However, at some age levels there was little difference between the correlations for Verbal Scale and Performance Scale with S-B IQ, and in a few instances the Performance Scale correlation was greater.

3. The correlation with Stanford-Binet IQ's will not be significantly different for WISC Full Scale and WISC Verbal Scale IQ's. Two studies report practically identical correlations between these two WISC scales and the Stanford-Binet [8, 9]. Weider, Noller, and Schramm [11] found identical correlations for their subjects as a whole, but when the group was divided with respect to age a difference appeared for the younger group ($r = .82$ for the Verbal Scale, $r = .90$ for the Full Scale). Pastovic and Guthrie [5] indicate approximately equal correlations with the Stanford-Binet for the two Scales in one study which they report but find differences of from .06 to .08 in three others. The difference between the correlations is .09 in the data of Frandsen and Higginson [1] and .18 for the total group of subjects studied by Krugman, Justman, and Wrightstone [2].

4. There is a significant relationship be-

---

[1] The writer wishes to express his appreciation to the following persons who did testing for this study: Joyce Beak, Annette Lawton, Patricia Marshall, Marie O'Hara, Edward Ormsby, Virginia Sandborg, Gordon Southon, and W. G. Willis.

tween the age of the subjects tested and differences in Stanford-Binet and WISC IQ's. Krugman, Justman, and Wrightstone [2] found a significant relationship between chronological age and differences between S-B and WISC IQ's for both the Verbal Scale and the Full Scale. A chi-square analysis indicated that the differences were greater for their younger subjects, with the WISC IQ's being lower. Pastovic and Guthrie state, "We conclude that the WISC should not be interpreted as equivalent to a Binet IQ at age levels below 10 years since the WISC score is consistently lower than that of the Binet" [5, p. 385]. Still more evidence comes from the work of Weider, Noller, and Schramm [11], who found significant differences in mean IQ between the Stanford-Binet and the WISC for subjects up to 7 years, 11 months old, but relatively small differences for older subjects. Frandsen and Higginson [1] report differences which are statistically significant for their group of fourth-graders ranging in chronological age from 9 years, 1 month to 10 years, 3 months. The three studies of defective subjects with mean chronological ages of 11 years, 11 months or older [4, 8, 9] have agreed in finding relatively small differences between mean S-B and WISC IQ's with the WISC IQ's being consistently higher. Nale [4] reports, however, that the difference of 2.59 IQ points which he found is significant at the .01 level of confidence.

5. There is a significant relationship between the size of Stanford-Binet IQ's and the difference between Stanford-Binet and WISC IQ's. Such a relationship is reported by Krugman, Justman, and Wrightstone [2], who correlated S-B IQ's with the differences between S-B and WISC IQ's for all three of the WISC Scales. The three product-moment correlations obtained were all significantly greater than zero.

Although not pertinent to the comparison of Stanford-Binet and WISC IQ's, two additional hypotheses of some interest are suggested by the studies done to date.

6. There is a significant difference between mean WISC Verbal Scale and Performance Scale IQ's. This is the contention of Pastovic and Guthrie who state, ". . . our data and those of three other studies cited, find a higher mean performance IQ than verbal IQ over a wide range of intelligence, although the WISC was intended to have a difference of zero points between these averages" [5, p. 385]. Seashore [7], however, on the basis of the test data for the entire WISC standardization sample, found zero difference between the mean IQ's of the Verbal and Performance Scales.

7. There is a significant correlation between WISC Full Scale IQ's and the difference between WISC Verbal Scale and Performance Scale IQ's. The evidence on this point is weak, being based on the observation of defective subjects only. Seashore [7] reports that for 55 feebleminded subjects in the WISC standardization sample the mean "discrepancy score" (Verbal Scale IQ minus Performance Scale IQ) was a -2 points. A difference in mean IQ of 4.9 points (Performance Scale mean IQ greater) was found by Sloan and Schneider [8]. A difference in the same direction of 5 IQ points is reported by Stacey and Levin [9]. For both of these latter comparisons the difference in means is statistically significant at the .001 level of confidence (computed by this writer). Whether subjects of superior intelligence have higher Verbal Scale than Performance Scale IQ's has not yet been reported, but the possibility seems worth investigating.

Procedure and Subjects

The testing for the present study was done by a small group of senior and graduate students of the University of California at Los Angeles. These students had completed previous to the study or were in process of completing concurrently at least a year's work in intelligence testing, including a semester of intelligence test practice under the supervision of the writer.

These students were instructed to examine at least ten subjects of varying ages with both the Stanford-Binet Intelligence Scale and the Wechsler Intelligence Scale for Children. Half of their subjects were to be given the Stanford-Binet first and half the WISC first. The test booklets were then checked by the writer for scoring errors and evidences of errors in administration. From the test booklets which remained after this screening, 26 were chosen representing subjects who were given the WISC first. The ages of the subjects ranged from 5 through 13 years with a mean of 9.3 years and a standard deviation of 2.5 years. Seventeen were boys and 9 were girls. The mean Stanford-Binet IQ for this group was 113.7 (*SD* 15.8). An additional 26 test booklets were then chosen representing subjects who had received the Stanford-Binet first, the selection being made in such a way as to reproduce so far as possible the statistical characteristics of the group mentioned above. For the second group the mean chronological age was also 9.3 years and the *SD* 2.5 years. Fourteen of this group were boys and 12 were girls. The mean Stanford-Binet IQ was 113.7 (*SD* 15.5). After the equated groups were made up, an additional 20 records remained representing subjects who had been tested with the S-B first. These were added to those mentioned above for the purposes of the chi-square analyses to be reported.

The longest interval between the two tests was 84 days (1 subject), and the shortest interval was 2 days (2 subjects). Most subjects were given the two tests a week apart, the median interval for the entire group being 7.5 days. Forty-one of the subjects were tested with Form L of the Stanford-Binet and 11 were tested with Form M.

Results

1. *The effect of order of testing on WISC IQ's.* For the group examined with the WISC first the mean IQ's obtained were 110.4 for the Verbal Scale, 110.8 for the Performance Scale, and 111.4 for the Full Scale. For the group examined with the Stanford-Binet first the corresponding mean IQ's were 111.3, 111.3, and 112.5. The *t* values [6] for the differences between means were .22 for the Verbal Scale, .13 for the Performance Scale, and .27 for the Full Scale. None of the differences are statistically significant.

2. and 3. *The correlation of Stanford-Binet and WISC IQ's.* For the 52 subjects in the equated groups the product-moment correlation of the Stanford-Binet with WISC IQ's was .88 for the Verbal Scale, .73 for the Performance Scale, and .87 for the Full Scale. The critical ratio for the difference between the first two correlations, following the statistical procedure suggested by McNemar [3], is 2.96, and the probability that the correlations are equal is less than .002. The critical ratio for the differences between the second and third correlations is 4.55, indicating a probability of less than .001 that the correlations are equal. However, when the correlations of the Stanford-Binet with the Verbal Scale and with the Full Scale are compared, the critical ratio for the difference in correlations is only .36. It is obvious that for this comparison the null hypothesis cannot be rejected with any appreciable degree of confidence.

4. *Age and differences between Stanford-Binet and WISC IQ's.* The statistical analysis of the relation between age and differences between Stanford-Binet and WISC IQ's is based on the records of 72 subjects. A 2 × 2 table was set up dividing subjects with respect to age at the 10-year level and with respect to differences between Stanford-Binet and WISC IQ's into those for which the difference was 5 points or less and those for which the difference was 6 points or greater. The value of chi square for differences between Stanford-Binet and Verbal Scale IQ's was 1.006, which has a probability greater than .30 of occurring in samples in which there is no relationship between the variables compared. The value of chi square for Stanford-Binet and Performance Scale IQ differences was .082, *p* greater than .70. The value of chi square for differences between Stanford-Binet and Full Scale IQ's was 1.47, *p* greater than .20. None of the chi-square values warrants the rejection of the null hypothesis of no relationship between chronological age and differences in Stanford-Binet and WISC IQ's.

To provide data comparable with those of other studies, mean IQ's were computed separately for subjects 10 years and older and for subjects less than 10 years old. The mean Stanford-Binet IQ for 23 subjects 10 years or older was 115.3. The mean WISC IQ's were 113.2 for the Verbal Scale, 112.9 for the Performance Scale, and 113.6 for the Full Scale. The *t* test [6] indicates that none of the mean WISC IQ's can be considered significantly different from the mean Stanford-

Binet IQ for these subjects, the t values being 1.35 for the Verbal Scale, 1.00 for the Performance Scale, and 1.00 for the Full Scale. The mean Stanford-Binet IQ for the 29 subjects in this study who were less than 10 years old was 112.3. The mean WISC IQ's were 109.1 for the Verbal Scale, 109.6 for the Performance Scale, and 110.6 for the Full Scale. The corresponding t values for differences between mean Stanford-Binet and mean WISC IQ's are 2.26 for the Verbal Scale, 1.37 for the Performance Scale, and 1.41 for the Full Scale. Of these t values, only the first approaches statistical significance, the probability being .012 of obtaining a difference as large as that observed when there is no real difference in the means compared.

5. *Stanford-Binet IQ's and differences between Stanford-Binet and WISC IQ's.* The records of 72 subjects were used for the chi-square analysis of the relationship between Stanford-Binet IQ's and the difference in Stanford-Binet and WISC IQ's. Subjects were divided into those with Stanford-Binet IQ's of 114 or greater and those with Stanford-Binet IQ's of less than 114. Each group was then divided into those with Stanford-Binet and WISC IQ differences of 6 points or more and those with differences of 5 points or less. The value of chi square when Verbal Scale IQ's were involved was .845 with a p greater than .30 favoring the null hypothesis. Chi square when Performance Scale IQ's were involved was .716 with p greater than .30. Chi square when Full Scale IQ's were involved was .258 with p greater than .50. None of the chi-square values indicates a significant relationship between Stanford-Binet IQ's and the difference between Stanford-Binet and WISC IQ's.

6. *The difference between WISC Verbal Scale and Performance Scale IQ's.* The data for 52 subjects in the present study show a mean difference of 0.0 between Verbal Scale and Performance Scale IQ's on the WISC. The standard deviation of the distribution of differences was 10.15 IQ points. The standard error of the mean difference is 1.42, which would indicate that for the universe represented by this sample the probability is .01 that the mean difference might be greater than 3.75, either plus or minus.

When the differences between Verbal and Performance Scale IQ's are taken without regard to algebraic sign the data of the present sample indicate a crude mode for the distribution of 4.5 points difference in IQ with 27% of all differences being either 4 or 5 points. The median absolute difference was 6.8 points and the mean absolute difference 8.4 points. Q_1 for the distribution of absolute differences was 4.43, and Q_3 was 12.0.

7. *WISC Full Scale IQ's and the difference between Verbal and Performance Scale IQ's.* For 52 subjects the discrepancy scores (Verbal Scale IQ minus Performance Scale IQ) suggested by Seashore [6] were computed and correlated with WISC Full Scale IQ's. The product-moment correlation was .14 with a standard error of .14. Such a correlation cannot be considered as significantly different from zero.

Discussion

The present study fails to support many of the trends noted in earlier studies of the WISC. Although several investigators have alternated the order of testing with the WISC and Stanford-Binet (and there is no reason for failing to do so in the future as a matter of precaution), the present report indicates that the order of testing does not have a significant effect on mean IQ for any of the WISC Scales. The data of this study agree with those of the majority of previous investigations in finding that the Stanford-Binet does not correlate as highly with the Performance Scale of the WISC as it does with the Verbal and Full Scales. They also indicate no significant difference in the correlation of the Stanford-Binet with the Verbal Scale and with the Full Scale IQ's of the WISC, although previous information on this point has varied among different investigators.

Contrary to the conclusions of at least two other investigations [2, 5] the present study fails to find clear-cut evidence of a relationship between the age of subjects tested and the difference in Stanford-Binet and WISC IQ's. A chi-square analysis of absolute differences and age failed to show a significant relationship. Nor was it possible to find a significant relationship between the size of Stanford-

Binet IQ's and the amount of difference between Stanford-Binet and WISC IQ's.

The present study agrees with Seashore's [7] report on the standardization sample of zero mean difference between Verbal and Performance Scale IQ's and suggests that this difference cannot in all probability be greater than 3.75 IQ points in the universe represented by the sample studied. The tentative hypothesis that there might be a relationship between WISC Full Scale IQ and differences between Verbal and Performance Scale IQ for normal subjects was not substantiated.

In considering the results of the studies of the WISC to date, the writer wishes to put forth some tentative suggestions. The first is that the use of the WISC with high-grade defectives appears to give results which are relatively unique to the defective group. (This is illustrated by the fact that for all reports in this area the mean Performance Scale IQ is consistently greater than mean Verbal Scale IQ.) The second is that generalizations about the WISC based upon the study of a relatively homogeneous age group may not apply to groups of subjects which are more heterogeneous with respect to age.

The question remains as to why the results of the present study appear to differ from those of the two studies employing normal subjects of various ages [2, 11]. So far as the study of Weider, Noller, and Schramm [11] is concerned, the present data agree on the extent of correlation between Stanford-Binet and WISC IQ's but differ with respect to the relation of chronological age to differences between Stanford-Binet and WISC IQ's. The subjects used in the earlier study were somewhat younger than those in this study and had lower mean IQ's. However, even when only subjects less than 8 years old from the present study are considered (to form a group comparable to the "younger" group in the previous study), there is still no significant difference between Stanford-Binet and WISC IQ's. Neither was it possible to find a significant relationship between Stanford-Binet Mental Ages and Stanford-Binet—WISC IQ differences. The only remaining obvious difference in the nature of the two studies lies in the fact that the Stanford-Binet and WISC were usually given either in two sessions on the same day or in a single session in the Weider, Noller, and Schramm study. This suggests the possibility of constant errors of measurement (particularly with younger subjects) based on fatigue, loss of interest, and similar progressive changes in the subject. There is no indication in the report that the order of testing was varied, which might become an important consideration when the interval between tests is quite short.

Even this possibility does not exist to account for the differences in results between the present study and that of Krugman, Justman, and Wrightstone [2]. Their work was much the more extensive, including a total of 332 subjects well distributed throughout the age range to which the WISC is applicable. It must be considered the definitive study of Stanford-Binet—WISC comparisons to date. Nevertheless, the fact that the present study failed to find a significant relationship between either chronological age or Stanford-Binet IQ and the difference between S-B and WISC IQ's may serve to stimulate further research and to suggest caution in making generalizations.

Since the fact that student examiners gave the scales in the present study might suggest that this was a factor in the results obtained, the writer has compared the results of their testing with that of the standardization testing on two points where comparable data are available. As noted above, Seashore [7] found a mean difference of zero between Verbal and Performance Scale IQ's and a standard deviation for the differences of 12.5 points. He reports a median of 8 points for the absolute differences. The data of this study show a mean difference of zero points with a standard deviation for the differences of 10.2 points and a median of 6.8 points for the absolute differences. For his subjects 10½ years old Wechsler [10] reports correlations of .68 between Verbal and Performance Scale IQ's, .93 between Verbal and Full Scales, and .90 between Performance and Full Scales. For the present study (in which the mean age was 9.3 years) the corresponding correlations are .75, .94, and .92.

Summary

Seventy-two subjects were given both the Wechsler Intelligence Scale for Children and the Stanford-Binet Revised Intelligence Scale. Twenty-six subjects were examined with the WISC first and could be equated, as a group, with 26 subjects examined with the Stanford-Binet first. A statistical analysis of the results indicated:

1. No significant practice effect on WISC IQ's when the Stanford-Binet was given first and the median interval between tests was 7 days.
2. Significant differences between the correlation of Stanford-Binet with Performance Scale IQ's and the correlations of Stanford-Binet with either Verbal or Full Scale IQ's on the WISC.
3. No significance for the difference between the correlation of the Stanford-Binet with Verbal Scale and its correlation with Full Scale IQ's on the WISC.
4. No significant relationship between chronological age and the differences between Stanford-Binet and WISC IQ's for any of the WISC Scales.
5. No significant relationship between Stanford-Binet IQ and the differences between Stanford-Binet and WISC IQ's for any of the WISC Scales.
6. Zero mean difference between WISC Verbal Scale and Performance Scale IQ's with a standard error of the mean of 1.42 points.
7. No significant relationship between WISC Full Scale IQ's and the difference between WISC Verbal Scale and Performance Scale IQ's.

Received August 19, 1952.

References

1. Frandsen, A. N., & Higginson, J. B. The Stanford-Binet and the Wechsler Intelligence Scale for Children. *J. consult. Psychol.*, 1951, **15**, 236-238.
2. Krugman, Judith I., Justman, J., & Wrightstone, J. W. Pupil functioning on the Stanford-Binet and the Wechsler Intelligence Scale for Children. *J. consult. Psychol.*, 1951, **15**, 475-483.
3. McNemar, Q. *Psychological statistics.* New York: Wiley, 1949.
4. Nale, S. The Childrens-Wechsler and the Binet on 104 mental defectives at the Polk State School. *Amer. J. ment. Def.*, 1951, **56**, 419-423.
5. Pastovic, J. J., & Guthrie, G. M. Some evidence on the validity of the WISC. *J. consult. Psychol.*, 1951, **15**, 385-386.
6. Peatman, J. G. *Descriptive and sampling statistics.* New York: Harper, 1947.
7. Seashore, H. G. Differences between Verbal and Performance IQs on the Wechsler Intelligence Scale for Children. *J. consult. Psychol.*, 1951, **15**, 62-67.
8. Sloan, W., & Schneider, B. A study of the Wechsler Intelligence Scale for Children with mental defectives. *Amer. J. ment. Def.*, 1951, **55**, 573-575.
9. Stacey, C. L., & Levin, Janice. Correlation analysis of subnormal subjects on the Stanford-Binet and Wechsler Intelligence Scale for Children. *Amer. J. ment. Def.*, 1951, **55**, 590-597.
10. Wechsler, D. *Wechsler Intelligence Scale for Children.* New York: Psychological Corp., 1949.
11. Weider, A., Noller, P. A., & Schramm, T. A. The Wechsler Intelligence Scale for Children and the Revised Stanford-Binet. *J. consult. Psychol.*, 1951, **15**, 330-333.

Journal of Consulting Psychology
Vol. 17, No. 2, 1953

Note on Elwood's Study of IQ Changes

Quinn McNemar
Stanford University

A recent paper by Elwood [1] purports "to examine the validity" of corrections, suggested by the present writer [2, pp. 172-174], to IQ's for differences in IQ variability at different age levels. She reasons that since corrections at age 6 tend to make low IQ's lower, one would expect those with low (uncorrected) IQ's at age 6 to have lower values at age 8, where no corrections are needed.

Her method involves a comparison of test-retest IQ's of three groups of children tested and retested at about ages 6 and 8, the three groups having been *selected* on the basis of the first test as those falling in the three intervals 50-59, 60-69, and 70-79 on the IQ scale. Then on the basis of averages, the "expected" drop in IQ's at the second testing did not occur, hence it is implied that the recommended corrections are not sound.

What Elwood failed to consider is the fact that a regression upward has neatly counterbalanced the "expected" drop in IQ. Indeed, the lower variability at age 6, compared to that at age 8, is responsible for the failure of these cases, so chosen, to show a regressive *gain*. If Elwood wishes to make an unambiguous factual contribution to this problem by the longitudinal approach, she should simply compare the *SD* for *all* cases tested at circa age 6 with the *SD* for the same cases retested two years later. It must, of course, be presumed that the cases are unselected, i.e., typical of 6-year-olds.

Received August 27, 1952.

References

1. Elwood, Mary I. Changes in Stanford-Binet IQ of retarded six-year-olds. *J. consult. Psychol.*, 1952, **16**, 217-219.
2. McNemar, Q. *The revision of the Stanford-Binet Scale.* Boston: Houghton Mifflin, 1942.

NEW BOOKS AND TESTS

Books

Bowers, Henry. *Research in the training of teachers.* Toronto: J. M. Dent & Sons (Canada) Ltd., & Macmillan Co. of Canada, 1952. Pp. vii + 167. $1.90.

The book consists of a series of fourteen papers dealing with some of the many factors that influence a student's success in practice teaching. Each paper is a separate unit dealing with one phase of the problem. They are carefully organized, stating clearly the question, the methods used, the results, and the conclusions obtained from the study. The concluding summary analyzes the student's background, personality traits, interests and activities, social acceptance, capacity for leadership and other factors, in relation to their effect on the student's success in practice teaching. The book should be of special interest to those who are training our future teachers, and to high-school counselors who are advising students in regard to their possible success as teachers. One wishes that the study were based on success in teaching after several years of work, rather than on the more academic situations found in practice teaching.—*B. M. L.*

Davidson, Henry A. *Forensic psychiatry.* New York: Ronald Press, 1952. Pp. viii + 398. $8.00.

This book is designed as a psychiatric-legal guide for physicians, particularly psychiatrists, who may be called upon as expert witnesses. It begins with a careful discussion of the McNaghten formula, the basis for the determination of criminal responsibility in this country. It includes description of types of examination and reports, and discussion of the physician's role in cases involving personal injury, annulment of a marriage, custody of children, evaluation of testamentary capacity, appraisal of sex offenders, alcoholics, juvenile offenders and malingerers, and a description of commitment procedures and the competency of the mentally ill. The situations chosen to illustrate the discussion of malpractice as it applies to the psychiatrist will be of interest to psychologists, particularly one in which a psychologist was the chief target of criticism. The book also includes a discussion of courtroom practice. — *A. R.*

Eissler, Ruth S. *et al. The psychoanalytic study of the child, Vol. VII.* New York: International Universities Press, 1952. Pp. 448. $7.50.

Of 21 chapters in this volume, the first 4 constitute a symposium on the Development of Ego and Id, held at the Congress of the International Psychoanalytical Association at Amsterdam, Holland, in 1951. This symposium is highly technical, and probably of interest only to analysts. Most psychologists will probably find the other chapters of greater interest, since they are more clinically oriented. The majority of them deal with very young children—age two to four—and three sections deal with anxiety factors related to hospitalization of young children. An interesting methodological approach in the second of these three discussions is the use of a film for the psychoanalytic study of a young child. One of the most valuable discussions of this volume is a 65-page chapter by Gerald H. J. Pearson, "A Survey of Learning Difficulties in Children." This is both a survey of the psychoanalytic literature on the subject, as well as an elaboration of the author's thesis, "The ability to learn is a function of the ego and therefore its disorder is produced by influences which affect the ego functions." Seven major factors causing learning difficulties are discussed, and abstracts from 37 cases are used illustratively.—*M. K.*

Eysenck, H. J. *The scientific study of personality.* London: Routledge and Kegan Paul, and New York: Macmillan, 1952. Pp. xiii + 320. $4.50.

This volume, in a sense a sequel to the *Dimensions of Personality* (1947), reports more of Eysenck's original and ingenious experiments. His use of objective tests in the appraisal of personality, in contrast to questionnaires and projective methods, deserves wider attention in the United States. The studies lead to one main proposition, of the existence of three dimensions of personality which are relatively independent: introversion-extroversion, neuroticism, and psychoticism. In this frame of reference ". . . the question: 'Is this person psychotic or neurotic?' becomes as unreasonable as the question: 'Is this person intelligent or tall?' " (p. 285). The experiments which support these hypotheses are, in the main, neatly designed and analyzed, and yield sharp tests of propositions deduced from theory. Some of them, however, seem to suffer from the lack of cross validation on independent groups. Such a statically cross-sectional approach, however ably ap-

---

Note: The reviews were prepared by the Editor and the Associate Editors, who may be identified by their initials.

plied, leads almost inevitably to a hereditarian position. Eysenck espouses "constitution" as the cause of individual differences in personality, and generally belittles psychogenic hypotheses based on clinical evidence. One may hope that someday he will undertake the experimental investigation of learning, and apply his considerable talents to the study of change as well as to the study of status. —*L.F.S.*

Fiedler, Miriam F. *Deaf children in a hearing world.* New York: Ronald Press, 1952. Pp. viii + 320. $5.00.

The notion that handicapped children should be reared and educated in close association with normal children is widely accepted but all too rarely practiced. This semipopular book, for parents and teachers as much as for psychologists and other specialists, reports experiences arising from an experiment in the normal education of deaf and hard-of-hearing children. The chapters contain an account of the experiment, case studies of children, and discussions of parents' problems based on transcribed recordings of discussion groups.—*L.F.S.*

Gilbert, Jeanne G. *Understanding old age.* New York: Ronald Press, 1952. Pp. ix + 422. $5.00.

This is a welcome addition to the growing literature on senescence, by a pioneer, productive, well-qualified writer. As an organized informational compendium it will be helpful both to present workers in the field and to those who may not yet be versed in the field's importance. It leaves unsolved many issues which becloud the critical evaluation of the many studies here collected. Is gerontology "the scientific study of the phenomena of aging" (opening sentence, p. 3)? Is aging a continuing anabolic-catabolic reciprocation from conception to death, or is it some later or final segment of growth and decline? Which of its instar stages is really *old* age? And do we "age" at equal rates at all life-span periods, or do our organ-systems regress unequally during involution as they progressed in evolution? Who is old, when, why, and by what individual differences? Is longevity positively correlated with favorable aptitudes, or do the good die young? What constitutes a standard sample for experimental and control studies? Does "old age" comprise a "clinical" syndrome? If so, is it primarily psychosomatic or somatopsychic? Until such issues are resolved one may still endorse the hackneyed quip that "old age is any age older than mine." Dr. Gilbert recognizes these problems and offers much material toward their solution. Adding to our knowledge about old age increases our understanding of it. — *E.A.D.*

Hilgard, Ernest R., Kubie, Lawrence S., & Pumpian-Mindlin, E. *Psychoanalysis as science.* Stanford, Calif: Stanford Univer. Press, 1952. Pp. x + 174. $4.25.

The title of this book is eye-catching, but its content turns out to be something of a disappointment. It is a clear and able summary of much that has been said about psychoanalysis, but contains no strikingly new evidence. Of five lectures given at California Institute of Technology in 1950, the first two, by Hilgard, summarize the findings of experimental psychology which are relevant to psychoanalytic dynamics and therapy. Two lectures by Kubie describe the techniques of psychoanalysis, with persuasive but nonexperimental illustrative anecdotes, and cite the main premises of psychoanalysis. In the final lecture, Pumpian-Mindlin traces the historical development of psychoanalysis and its relationships to the biological and social sciences.—*L.F.S.*

Krieg, Wendell J. L. *Functional neuroanatomy* (2nd Ed.). Philadelphia: Blakiston, 1953. Pp. xviii + 659. $9.00.

This magnificent text and book of charts deserves to be brought to the attention of all psychologists. It is outstanding in two respects. First, it is more genuinely functional than most books on neuroanatomy with clear descriptions of what the neural structures do. Second, many of its 274 figures, and its 88 pages of atlas, have a remarkable three-dimensional quality which facilitates visualizing the nervous system "in the round." While most effectively used in connection with a laboratory course, the text is a suitable adjunct for any course in physiological psychology.—*L. F. S.*

Reik, Theodor. *The secret self.* New York: Farrar, Straus, & Young, 1952. Pp. 329. $3.50.

In his fiction-like style Reik opens with a conversation he had with Freud, and closes with satisfying insights he experienced one dawn as he ended an insomnious night of personal struggle with deep sleep. Such is the tone of another volume devoted to the thesis that psychoanalysis is a great adventure in inner experience and not a routine with mental gadgets manipulated by skilled technicians of the unconscious. To be a creative experience for the patient it must first be such for the analyst. And so it is for Reik. He demonstrates this in twenty chapters of subjective notes linked with case material and classical literature. These are "impressions obtained in everyday life and in analytic sessions [which] awakened echoes of other inner experiences shaped by a great writer." Reik adds a contribution indeed to our contemporary scientific emphasis in clinical psychology. His voice reminds us that the direction of human adjustment is at best a creative process, an aspect which needs emphasis in our training program. Science and this subjective emphasis together may teach us that the analyst can "listen to his patient and to his own associations," but this art is a highly finished one requiring much nurture and difficult to validate. — *F. McK.*

Shaffer, G. Wilson, & Lazarus, Richard S. *Fundamental concepts in clinical psychology.* New York: McGraw-Hill, 1952. Pp. xi + 540. $6.00.

This unusually comprehensive text on the theory and practice of clinical psychology is written clearly and well. While suitable for advanced undergraduates and beginning graduate students, more advanced students may read it with profit. At many points, the mature practicing clinician may wish for a more penetrating discussion, but the book will offer him an integrated view of the field that is not easily obtainable elsewhere. The treatment ranges well beyond the narrow field of clinical psychology itself to supply supplementary background material from all of psychology. Chapters on scientific methodology and on the nature of personality illustrate this breadth. While this comprehensive approach inevitably necessitates some superficiality of treatment at times, the net result remains at a high level, and the book should make an excellent text. — *W. A. H.*

Wechsler, David. *The range of human capacities.* Baltimore: Williams & Wilkins, 1952. Pp. ix + 190. $4.00.

The changes in this revision consist primarily in the addition of two chapters, on span of life as a human capacity, and on range in productive operations, together with increased emphasis on the reality of decreased capacity with aging. The discussion is based on the concept of the validity of a range ratio as a measure of human capacity. The ratio settled upon is that between the 2nd and 999th individual in every thousand. After examination of a quantity of data, including 26 new sets of measurements, the author believes that "human variability, when compared to that of other phenomena in nature is extremely limited, and the differences which separate human beings from one another with respect to whatsoever trait or ability we may wish to compare, are far smaller than is ordinarily supposed." The other phenomena in nature to which he refers are various physical constants; no biological data other than those on man are adduced. Omitting special cases, the total range ratios presented fall between 1.16:1 and 2.93:1. Their distribution is multimodal, indicating the following groupings: linear body measures (ave. 1.30:1), metabolic rates (1.39:1), body circumference (1.52:1), physiologic function (2.07:1), motor coordination and speed of movement (2.23:1), measures of body weight (2.33:1), and perceptual and intellectual abilities (2.58:1). — *A.R.*

Tests

Bellak, Leopold, & Bellak, Sonya S. *Supplement to the Children's Apperception Test (CAT-S).* Ages 3–10. Individual test. 10 pictures, with manual, pp. 8 ($6.00). New York (Box 42, Gracie Station): C. P. S. Co., 1952.

The ten supplementary pictures for the CAT are designed for use with children whose complaint problems or previous responses suggest particular areas of investigation. The ten situations, enacted by human-like animals comparable to those in the CAT, represent peer play, schoolroom, playing house, mothering, injury, competition, body image, visit to physician, bathroom scene, and pregnant mother. The manual recommends several methods for using the pictures, and describes typical responses, with limited normative data on 40 six- and seven-year-olds. These ingenious materials will be a useful tool for research and clinical appraisal. — *L. F. S.*

Jay, Edith Sherman. *A book about me.* Preschool-gr. 1, ages 4–7. 1 form. Untimed. Booklets ($2.40 per 10); analysis sheets (55¢ per 25); manual, pp. 32 (25¢); specimen set (50¢). Chicago: Science Research Associates, 1952.

A Book About Me provides material for a survey of children's backgrounds, maturity, activities, and interests, at the prekindergarten to first-grade level. Children mark pictures in an attractive book according to standard instructions for a formal survey, or according to more flexible procedures for other uses which may be planned by the teacher or investigator. The manual describes the survey procedures and the auxiliary analysis sheet, and suggests further uses of the material in school activity programs. — *L. F. S.*

Books Received

Frankel, George W. *Let's hear it.* New York: Stratford House, 1952. Pp. 63. $1.00.

Garrett, James F. (Ed.) *Psychological aspects of physical disability.* Federal Security Agency, Office of Vocational Rehabilitation, Rehabilitation Service Series No. 210. Washington: U.S. Government Printing Office, 1952. Pp. vii + 195. 45¢.

Rickman, John (Ed.) *On the bringing up of children.* (2nd Ed.) New York: Robert Brunner, 1952. Pp. xxii + 243. $3.00.

Steiner, Lee R. *A practical guide for troubled people.* New York: Greenberg, 1952. Pp. 209. $3.50.

Thompson, Charles B., & Sill, Alfreda P. *Our common neuroses.* New York: Exposition Press, 1952. Pp. xxxii + 210. $3.50.

Journal of Consulting Psychology
Vol. 17, No. 3, 1953

Process and Reactive Schizophrenia[1]

Robert E. Kantor
Palo Alto, California

Julius M. Wallner
Veterans Administration Hospital, Palo Alto

and C. L. Winder
Stanford University

The process-reactive hypothesis regarding schizophrenia may be stated as follows: Within the category schizophrenia two subgroups can be differentiated, these being the process or true or chronic, and the reactive or benign or acute. The hypothesis has usually been stated in the dichotomous form, as in this study, but may well develop into a dimensional concept as it is clarified empirically.

This hypothesis has appeared explicitly or implicitly in psychiatric and psychological literature since Bleuler [4, 5, 6] formulated the concept of schizophrenia in revision of Kraepelin's [14, 15] concept of dementia praecox. Bleuler recognized that there were some patients who recovered who were classifiable as dementia praecox according to Kraepelin's criteria, and, therefore, that further differentiation than that afforded by those criteria was necessary if maximum predictability was to be obtained. As Meduna and McCullough [22] have clearly argued, Bleuler's distinction has not been followed, and so the prediction of recovery among those diagnosed schizophrenic or dementia praecox has not been as effective as it might have been.

Kretschmer [16] and later Mauz [21] held one version of the process-reactive view, believing that true schizophrenia occurs preponderantly in those of asthenic or dysplastic body type. These authors thought that reactive schizophrenia occurs in those whose constitution does not render them prone to develop true schizophrenia.

Lewis [20] in his 1936 review of dementia-praecox researches arrived at the conclusion that investigation of the process-reactive hypothesis was a major indicated direction for further research.

Langfeldt [17] has published evidence in favor of the process-reactive view, finding that previously identified atypical cases of schizophrenia improved much more than typical cases. Langfeldt, in later papers [18, 19], has presented evidence that indefinite schizophrenia, with a favorable prognosis, can be differentiated from definite schizophrenia.

Milici [24] and Paskind and Brown [25] report cases diagnosed as schizophrenia in which the precipitating environmental stress or shock was known and where recovery or remission was relatively early and complete. Reports of many such cases were frequent during and after World War II and will not be cited. The inference based on these studies has usually been that there is a schizophreniform psychosis which develops suddenly and in response to relevant stress and which does not follow the classical course, and that many cases of so-called typical schizophrenia are so classified mistakenly.

Extensive studies of the childhood histories of schizophrenics with a view to implications for prognosis have been reported by Wittman [26, 27] and by Wittman and Steinberg [28, 29]. These studies carry on in the tradition of Meyer [23], as do studies of Hoch [8] and Kasanin and Veo [12]. The view that the shut-in personality and onset at early ages are characteristic of the process group is supported by the results of these studies. Similarly, Hunt and Appel [9] distinguished a group of patients whom they labeled schizoaffective and among whom poorer environmental adjustment prior to illness was associated with poor prognosis.

---

[1] From the Veterans Administration Hospital, Palo Alto, California.

Among investigators who have found that prognosis in schizophrenia is related to the nature of the onset, Kant [10, 11] is notable. Unrecovered cases were characterized by gradual onset or a unique precipitating cause, whereas recovered cases had developed in the presence of pertinent stress. Kant has also presented evidence that Bleuler's criterion of a clear sensorium differentiates the nonrecovering group from those with better prognoses. This conclusion is supported also by evidence from Langfeldt [17] and Hunt and Appel [9].

A somewhat different type of evidence is represented by a survey of the literature on hospital remission rates reported by Blair [3]. He found that about 40 per cent of diagnosed schizophrenics seem to remit spontaneously, while about 60 per cent of such cases tend to remain unimproved or deteriorate. It can be argued that such statistics support the process-reactive hypothesis.

References to additional studies pertinent to the process-reactive view will be found in Bellak's [1] recent summary volume in which the general conclusions are in support of this general point of view. Ebaugh and Benjamin [7] had previously identified four groups of

Table 1

Items Defining Frame of Reference for Case History Judgments.

Process Schizophrenia	Reactive Schizophrenia
Birth to the fifth year	
a. Early psychological trauma	*a.* Good psychological history
b. Physical illness—severe or long	*b.* Good physical health
c. Odd member of family	*c.* Normal member of family
Fifth year to adolescence	
a. Difficulties at school	*a.* Well adjusted at school
b. Family troubles paralleled with sudden changes in patient's behavior	*b.* Domestic troubles unaccompanied by behavior disruptions. Patient "had what it took."
c. Introverted behavior trends and interests	*c.* Extroverted behavior trends and interests
d. History of breakdown of social, physical, mental functioning	*d.* History of adequate social, physical, mental functioning
e. Pathological siblings	*e.* Normal siblings
f. Overprotective or rejecting mother. "Momism"	*f.* Normally protective, accepting mother
g. Rejecting father	*g.* Accepting father
Adolescence to adulthood	
a. Lack of heterosexuality	*a.* Heterosexual behavior
b. Insidious, gradual onset of psychosis without pertinent stress	*b.* Sudden onset of psychosis; stress present and pertinent. Later onset
c. Physical aggression	*c.* Verbal aggression
d. Poor response to treatment	*d.* Good response to treatment
e. Lengthy stay in hospital	*e.* Short course in hospital
Adulthood	
a. Massive paranoia	*a.* Minor paranoid trends
b. Little capacity for alcohol	*b.* Much capacity for alcohol
c. No manic-depressive component	*c.* Presence of manic-depressive component
d. Failure under adversity	*d.* Success despite adversity
e. Discrepancy between ability and achievement	*e.* Harmony between ability and achievement
f. Awareness of change in self	*f.* No sensation of change
g. Somatic delusions	*g.* Absence of somatic delusions
h. Clash between culture and environment	*h.* Harmony between culture and environment
i. Loss of decency (nudity, public masturbation, etc.).	*i.* Retention of decency

schizophrenics as defined by the course of the disturbance. Just how many types of courses there are in schizophrenia is thus left in doubt. The present paper uses a dichotomy for convenience.

It is apparent from the literature available, that there are some patients who present a symptom complex very similar to the traditional schizophrenia syndrome, but who to some significant degree have a more favorable prognosis. If one were to formulate a reactive schizophrenia syndrome, it would contain the following elements: (*a*) the prepsychotic personality was relatively normal, (*b*) the onset of the psychosis was sudden with logical precipitating factors present, and (*c*) the patient did not maintain a clear sensorium.

The present study stems most directly from the work of Benjamin [2], who has pointed out that the Rorschach test is useful in revealing thinking disorders and who finds that some patients diagnosed schizophrenic on the basis of psychiatric examination do not display thinking disorders or display such signs in slight degree. Signs of thinking disorders are "relatively independent of the stage of illness and the momentary clinical condition of the patient. They are, to be sure, found in almost all so-called deteriorated cases, but are also seen, sometimes, in less pronounced form, at very early stages, occasionally long before a clinical diagnosis has been made, as well as after a clinical recovery from a severe attack" [2, p. 67]. Thinking disorders may well be process symptoms according to Benjamin and the absence of such signs may well be indicative of reactive schizophrenia. In the present study, a nonpsychotic Rorschach diagnosis is similar to Benjamin's absence of thinking disorder, i.e., indicative of reactive schizophrenia, and a diagnosis of psychosis is similar to Benjamin's thinking-disorder diagnosis, i.e., indicative of process schizophrenia.

Problem

The study reported here is concerned with three questions which arise if one entertains the process-reactive hypothesis.

1. Do diagnoses based upon Rorschach examination alone label as nonpsychotic a portion of the population of mental patients who are diagnosed clinically as schizophrenic?

2. Can case histories of clinically diagnosed schizophrenics be differentiated into two categories: reactive and process? A search of the literature revealed criteria which were used as the frame of reference in rating cases as process or reactive from social service case histories. The criteria are given in Table 1.

3. Are those cases rated psychotic from the Rorschach classified as process on the basis of the case histories and those cases judged nonpsychotic from the Rorschach classified as reactive from the case histories?

Method

Subjects. Two successive samples of subjects were studied. The first consisted of all patients at the Palo Alto Veterans Hospital who were diagnosed schizophrenic and who were given Rorschach tests during 1948, a total of 180. The second sample is all of the schizophrenics (clinical diagnosis) given the Rorschach during the first six months of 1949 at the same hospital. There were 147 such patients, three of whom were used in the first sample, giving a second-sample total of 144.

Any patient who could be placed in any of the following categories was eliminated from the study: (*a*) diagnosis of latent schizophrenia (4 cases), (*b*) diagnosis made from a battery of tests rather than the Rorschach alone (56 cases), (*c*) less than seven main scorable Rorschach responses (29 cases), (*d*) diagnosis of schizophrenia complicated by a secondary diagnosis (9 cases), (*e*) change of clinical diagnosis before an arbitrary date, i.e., May 2, 1949 for the first sample and August 1, 1949 for the second sample (23 cases). In all, 108 first-sample patients and 95 second-sample cases were retained for rating.

Diagnostic procedures and ratings. The psychiatric diagnoses were established on the basis of data collected by all appropriate services of the hospital, including psychological examinations. In each case, diagnosis was decided at a staff conference including representatives of all relevant hospital services and a psychiatric consultant. The final diagnosis was reviewed by the chief of psychiatry of the hospital. These diagnoses did not enter into the study directly, but defined the population for study.

The Rorschach diagnoses were made on the basis of tests given by trainee psychologists each of whom had adequate previous training in administration of the test. Analysis of the protocol was carried out by the person testing with the assistance of an advanced trainee. The final formulation and the protocol on which it was based were then reviewed by a staff psychologist who had prior interpretation experience with more than five hundred Rorschachs. The analysis criteria used are those outlined by Klopfer and Kelley [13].

The two judges who rated the case histories for the first sample were the chief of psychiatry of the Palo Alto Veterans Hospital and a clinical psychologist from Stanford University. Each judge rated from the same history materials at different times. There was no discussion of the cases by the two judges and none of the psychological examination materials was available at the time of rating. Cases were catagorized: Process, Reactive, or Cannot Say.

The chief of psychiatry had reviewed each case previously, and so had seen psychological examination and history materials together at a time prior to the time of making the ratings for this study. The clinical psychologist judge had seen the complete set of materials on only a very few of the cases at a previous time. Both judges are familiar with the diagnostic habits (psychiatric and psychological) at the hospital. The reader must decide for himself just how these factors contaminate the results. It is the conviction of the authors that the results were determined little by these factors.

As of the writing of this paper, the second sample has been rated only by the clinical psychologist judge. Before making the ratings, he had contact with only two of the patients and their Rorschach records and had seen none of the history materials of patients used in the second sample.

Results

1. Diagnoses based on the Rorschach protocols label as nonpsychotic a large portion of the population diagnosed schizophrenic by conventional psychiatric procedures. This conclusion is supported by the facts that of the 108 patients from the first sample, 57 were classified psychotic and 51 nonpsychotic from the Rorschach alone, and of the 74 cases of the second sample which were rated process or reactive, 38 were labeled nonpsychotic and 36 psychotic on the basis of the Rorschach. These results are in agreement with and support the conclusion stated above. This result in isolation might tempt some to be extremely critical of the clinical diagnoses or the Rorschach or both, since there is so little agreement between the two types of diagnoses.

2. Case histories of clinically diagnosed schizophrenics can be differentiated reliably into two catagories: reactive and process. The relevant results are presented in Table 2, the data being from the first sample. Of the 108 case histories, both judges arrived at a rating in 86 cases, and were in agreement in 64 of these cases. A χ^2 test, based on the distribution of cases shown in Table 2, shows that

Table 2

Agreement between Judges Concerning Process and Reactive Schizophrenia*

Judge B	Judge A Process	 Reactive	Total
Process	47	13	60
Reactive	9	17	26
Total	56	30	86

* $\chi^2 = 13.40, \quad p < .001$

Table 3

Comparison of Judgments from the Case History and Rorschach

Rorschach	Judgment Reactive	 Process
Judge A (First sample)		
Nonpsychotic	21	19
Psychotic	10	37
$\chi^2 = 7.87, \quad .01 > p > .001$		
Judge B (First sample)		
Nonpsychotic	29	21
Psychotic	5	52
$\chi^2 = 27.55, \quad p < .001$		
Judge A and B (First sample)		
Nonpsychotic	15	13
Psychotic	2	34
$\chi^2 = 16.24, \quad p < .001$		
Judge B (Second sample)		
Nonpsychotic	16	22
Psychotic	5	31
$\chi^2 = 7.24, \quad .01 > p > .001$		

this agreement is greater than would be expected by chance.

3. Those patients diagnosed as schizophrenic who are rated reactive from the history are most often judged nonpsychotic from the Rorschach and those rated process from the histories are most often judged psychotic from the Rorschach. Table 3 contains the relevant data. The independent judgments made from the case histories and those made from the Rorschach examinations are in greater agreement than could be expected by chance. This is the case for each judge on all rated cases and for the cases from the first sample on which both judges agreed.

Discussion

The central problem under consideration in this paper is the status of the reactive-process conception of schizophrenia. It is clear that this idea is still in a crude form. It is equally clear that the general idea has persisted, and it is the opinion of the authors that persistence of the view in the light of critical clinical evaluation reflects some validity. Granting this, our problem becomes the clarification, on an empirical basis, of this view.

In this study, some questions relevant to the process-reactive view have been investigated empirically, with full realization that the implications of the findings would be in some doubt. The procedures employed in this study are subject to some criticisms, most of which have been mentioned. For these reasons the study must be viewed as preliminary in nature.

It was determined that diagnosed schizophrenics could be further separated into the catagories *process* and *reactive* with some reliability. No information is available at present to indicate with precision the meaning of these categories. One thought is that the reactive cases are either milder or earlier than process cases. If Benjamin is correct in his opinion that the presence or absence of indicators of thinking disorders is relatively independent of duration and degree of illness, and if, as our data suggest, the process cases show evidences of thinking disorders while reactive cases do not, then the equating of reactive with early or mild schizophrenia is at least insufficient and is possibly incorrect.

Benjamin's work led him to adopt the point of view that schizophrenics with thinking disorders as measured by the Rorschach are apt to follow the classical dementia praecox course, whereas those who do not show thinking disorders on the Rorschach tend to follow an atypical course with some degree of recovery. We have not presented evidence directly relevant to this point, but a brief follow-up of our patients does indicate a higher rate of discharge from the hospital among cases called reactive than among the process cases.

The judgments of the case histories made in the present study were based on premorbid personality, the nature of the illness to some extent, and on some knowledge of the course of the disorder. The latter information was a part of our histories since most of the cases had been hospitalized in the service; better samples of cases would be groups of first-admission cases. The judgments of the Rorschach performances in this study were based in part on criteria of thinking disorders, but also upon general personality structure and function. The agreement with the results of Benjamin [2] is found even though the methods are not strictly parallel. This suggests that the general conclusions can be held with some confidence.

A brief comment on the lack of agreement between the Rorschach and the clinical diagnoses is perhaps in order. Suffice it to say that such agreement could undoubtedly be systematically varied depending on the particular diagnostic approaches used. There is no reason to believe that the diagnostic approaches at the hospital where this study was made are atypical in any important ways. It would seem that the Rorschach alone, as conventionally used, would be an inadequate criterion on which to base a psychiatric diagnosis. It seems equally clear that the Rorschach provides a relatively economical way of adding important data regarding the particular variety of schizophrenia being dealt with.

An additional implication of the results of the present study, taken with other evidence, is that diagnosis as usually practiced is not sufficiently specific to give maximum differentiating information about schizophrenic patients.

Summary

Two samples of schizophrenic patients were studied. In each case, the patient was classified as reactive or process on the basis of the history

and psychotic or nonpsychotic on the basis of the Rorschach. The following results were obtained:

1. Schizophrenic cases can be reliably differentiated into process and reactive groups.
2. Those cases classified as reactive from histories tend to be called nonpsychotic from the Rorschach.
3. Those cases classified as process from histories tend to be called psychotic from the Rorschach.

These results support the view that the diagnostic category of schizophrenia can be legitimately elaborated to include the classifications reactive and process. There is evidence that process and reactive cases differ in their psychological-function characteristics.

On the basis of these results, further analytical study of the reactive-process hypothesis is indicated. The full implications of the agreements demonstrated can be derived only from these further studies.

Received October 24, 1952.

References

1. Bellak, L. *Dementia praecox.* New York: Grune & Stratton, 1948.
2. Benjamin, J. D. A method for distinguishing and evaluating formal thinking disorders in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univer. Calif. Press, 1946. Pp. 66–71.
3. Blair, D. Prognosis in schizophrenia. *J. ment. Sci.,* 1940, **86**, 378–477.
4. Bleuler, E. *Dementia praecox oder Gruppe der Schizophrenien.* Leipzig: Deuticke, 1911.
5. Bleuler, E. The physiogenic and psychogenic in schizophrenia. *Amer. J. Psychiat.,* 1930, **10**, 203–211.
6. Bleuler, E. *Textbook of psychiatry.* Trans. by A. A. Brill. New York: Macmillan, 1936.
7. Ebaugh, F. G., & Benjamin, J. D. Trauma and mental disorders. In L. Brahdy & S. Kahn (Eds.), *Trauma and disease.* Philadelphia: Lea & Febiger, 1941. Pp. 272–273.
8. Hoch, A. Constitutional factors in the dementia praecox group. *Rev. Neurol. Psychiat.,* 1910, 8, 463–474.
9. Hunt, R. C., & Appel, K. E. Prognosis in psychoses lying midway between schizophrenia and manic-depressive psychoses. *Amer. J. Psychiat.,* 1936, **93**, 313–339.
10. Kant, O. Differential diagnosis of schizophrenia in light of concepts of personality-stratification. *Amer. J. Psychiat.,* 1940, **97**, 342–357.
11. Kant, O. Problem of psychogenic precipitation in schizophrenia. *Psychiat. Quart.,* 1942, **16**, 341–350.
12. Kasanin, J., & Veo, L. A study of the school adjustment of children who later in life become psychotic. *Amer. J. Orthopsychiat.,* 1932, **2**, 212–227.
13. Klopfer, B., & Kelley, D. G. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1942.
14. Kraepelin, E. *Dementia praecox.* Trans. from 8th German Ed. of *Textbook of psychiatry.* Edinburgh: E. J. Livingstone, 1918.
15. Kraepelin, E. *Dementia praecox and paraphrenia.* Trans. by Mary Barclay. Edinburgh: E. S. Livingstone, 1919.
16. Kretschmer, E. *Physique and character.* New York: Harcourt, Brace, 1925.
17. Langfeldt, G. Prognosis in schizophrenia and factors influencing course of disease: Catamnestic study, including individual re-examinations in 1936 with some considerations regarding diagnosis, pathogenesis and therapy. *Acta Psychiat. Neurol., Supp.* 13, 1937. Pp. 1–228.
18. Langfeldt, G. Prognosis and diagnosis of schizophrenia. *Norsk. mag. f. laegevidensk.,* 1938, **99**, 589–609.
19. Langfeldt, G. The diagnosis of schizophrenia. *Amer. J. Psychiat.,* 1951, **108**, 123–125.
20. Lewis, N. D. C. *Research in dementia praecox.* New York: National Committee for Mental Hygiene, 1936.
21. Mauz, F. *Die Prognostik der endogenen Psychosen.* Leipzig: G. Thieme, 1930.
22. Meduna, L. J., & McCullough, W. S. The modern concept of schizophrenia. *Med. Clin. N. Amer.,* 1945, Chicago Number, 147–164.
23. Meyer, A. Fundamental conceptions of dementia praecox. *Brit. med. J.,* 1906, **2**, 757–760.
24. Milici, P. Postemotive schizophrenia. *Psychiat. Quart.,* 1939, **13**, 278–293.
25. Paskind, H. A., & Brown, M. Psychoses resembling schizophrenia occurring with emotional stress and ending in recovery. *Amer. J. Psychiat.,* 1940, **96**, 1379–1388.
26. Wittman, Phyllis. Scale for measuring prognosis in schizophrenic patients. *Elgin State Hosp. Papers,* 1941, **4**, 20–33.
27. Wittman, Phyllis. Diagnostic and prognostic significance of the shut-in personality type as a prodromal factor in schizophrenia. *J. clin. Psychol.,* 1948, **4**, 211–214.
28. Wittman, Phyllis, & Steinberg, D. L. Follow-up of objective evaluation. *Elgin State Hosp. Papers,* 1944, **5**, 216–227.
29. Wittman, Phyllis, & Steinberg, D. L. Study of prodromal factors in mental illness with special reference to schizophrenia. *Amer. J. Psychiat.,* 1944, **100**, 811–816.

Journal of Consulting Psychology
Vol. 17, No. 3, 1953

Prognostic Criteria in the Case Histories of Hospitalized Mental Patients[1]

G. R. Pascal,[2] C. H. Swensen[2]
The University of Tennessee

Dorothy A. Feldman, Mary E. Cole, and Jean Bayard
University of Pittsburgh

The purpose of the present study is to report the relationships between certain aspects of information obtainable from the records of hospitalized mental patients and the outcome of treatment. The feasibility of such a project, and the many objections which can be raised against such an approach to prognosis, are outside the province of this study. Suffice it to say that in the course of planning the present study over 500 published papers were read which used information from case records in an attempt to find prognostic criteria. In spite of this voluminous literature, however, there is a distressing dearth of definitive studies.

The data upon which this study is based are precarious. In the seven years covered by the study different psychiatrists and psychiatric social workers of varying conceptions and capacities contributed to the materials which make up the case histories. Admission policies have changed. Ideas of treatment have fluctuated. Similar criticisms apply to other similar studies. Nevertheless, this approach to the study of prognostic criteria continues to be popular and has, indeed, resulted in several generally accepted and repeatedly found prognostic criteria, e.g., duration of illness [3, 24, 36, 38, 50], presence of affect [51, 49, 31, 27], and acuteness of onset [26, 45, 46, 48]. It is our contention, however, that the methodology usually employed does not make maximum use of the data and, in addition, tends to obscure rather than define, the relationships which may exist between variables and outcome of treatment.

We may, to illustrate our point, consider a specific variable, *duration of illness,* which has been generally found to be related to outcome of treatment both for the schizophrenic [3, 4] and manic-depressive [43] disorders. The usual procedure is to relate duration of illness to outcome of treatment without consideration of other variables which might affect outcome of treatment. The general finding has been that the shorter the duration the greater the chance of the patient to recover from his illness. Such findings have led to the oft held belief that early treatment is the key to recovery in mental illness. There is some evidence, on the other hand, to indicate that duration of illness as a prognostic variable may be an artifact. Wittman [53], for instance, in her extensive studies of prognosis, found a number of process schizophrenics with unfavorable outcome to be in the short-duration group. In a study comprising 960 cases, she found 134 with short duration who did not improve. She feels that by the usual statistical methods employed, the patients with a longer duration of illness include a heavier weighting of process schizophrenics, which gives long duration its poor prognosis.

The need is clear, we think, for better-controlled studies of such variables as duration of illness, mode of onset, age of onset, marital status, precipitating stress, etc. If data obtained from case records are to be used in a search for prognostic criteria, then more valid methods of study should be employed. When a variable such as duration of illness is studied, other variables, known to be related to outcome of treatment, need to be held constant. Such a statement does not mean to imply that all the variables which may affect outcome of treatment can be held constant while a single variable is being studied. Sufficent knowledge to encompass such a possibility, if it were feasible, is not now available. We do wish to state, however, that a much clearer notion of the relationship between duration of illness and outcome of treatment can be obtained if all the other variables obtainable from a case record and known to affect outcome are held constant while the duration is varied.

---

[1] The authors acknowledge their indebtedness to Drs. H. Brosin, Director, and F. Weniger, Clinical Director of the Western Psychiatric Institute, University of Pittsburgh, for permission to use the records. Special thanks are due Miss Charlotte Battles, secretary, for assistance in data gathering and as secretarial overseer of the project.

[2] At the University of Pittsburgh when this research was done.

The present study, of which this paper is the first in a series, attempts a controlled study of prognostic criteria obtained from case-record material. Variables known to be significantly related to outcome of treatment are extracted from case records. Single variables are studied while other significant variables are, insofar as possible, held constant.

Method and Procedure

The Western Psychiatric Institute and Clinic (WPIC) was opened in November, 1942. The records of all patients admitted to the hospital who were discharged not later than July, 1950 (to allow for a minimum of one year follow-up) were examined. These comprised 1118 records. The following selective criteria were now applied to the records:

age—15 to 55
lack of known cortical damage
absence of mental deficiency
absence of incapacitating somatic illness
absence of previous attacks of mental illness.

Table 1
Total Population: Status One Year after Discharge
According to Diagnostic Category

Diagnostic Category	N	Sex		Mean age	Mean ed. (in grades)	Status 1 yr. after disch.	
		M	F			Imp.	Unimp.
Schizophrenics							
Paranoid	123	67	56	33.2	10.6	36	87
Catatonic	76	35	41	24.8	10.8	38	38
Hebephrenic	32	19	13	25.9	9.8	3	29
Simple	11	7	4	26.8	13.2	6	5
Mixed	22	7	15	29.3	11.3	8	14
Total schizophrenic	264	135	129	29.3	10.7	91	173
Manic-depressives							
Manic	18	6	12	28.0	11.7	17	1
Depressed	19	10	9	36.6	10.0	10	9
Total manic-depressive	37	16	21	32.4	10.8	27	10
Involutional melancholia	21	3	18	44.3	7.8	14	7
Total affective psychoses	58	19	39	36.7	9.7	41	17
Paranoid condition	14	8	6	39.7	11.1	4	10
Other psychotics	3	2	1	32.3	10.7	2	1
All psychotics	339	164	175	31.0	10.5	138	201
Psychoneurotics							
Obsessive-compulsive	20	9	11	32.6	11.0	15	5
Hysteria	11	3	8	27.2	10.4	6	5
Anxiety	9	5	4	29.7	11.7	7	2
Reactive depression	20	7	13	34.0	11.5	17	3
Hypochondriasis	10	6	4	37.8	8.4	7	3
Mixed	25	10	15	33.6	9.7	19	6
All psychoneuroses	95	40	55	32.8	10.5	71	24
Behavior disorders	52	25	27	23.4	9.6	25	27
All nonpsychotic	147	65	82	29.5	10.2	96	51
All patients	486	229	257	30.5	10.4	234	252

The last criterion requires some explanation. The WPIC is primarily a referring hospital. Most of its patients come from other hospitals. The criterion "absence of previous attacks of mental illness" does not, therefore, mean that the patient was first hospitalized at WPIC. It means that upon admission to WPIC the patient was suffering from one continuous attack of mental illness without known prior attacks which resulted in referral to a psychiatrist. Application of these criteria left 486 cases available for further study.

Table 1 shows the composition of the 486 cases. The table shows the sex, mean age, mean education in grade years, and status one year after discharge by diagnostic category. Our interest, at this point of discussion, centers about the representativeness of the sample with respect to outcome of treatment. The sample contains 264 schizophrenics. Of these, 91 or 34 per cent improved. This figure compares favorably with that of Taylor and Van Salzen [48] whose study contained 1100 schizophrenic patients, 37 per cent of whom were able to complete, successfully, one year of parole. Within the schizophrenic group, only the catatonics and paranoids are in sufficient number to relate to other studies. Of the 123 paranoids, 36 or 29 per cent improved. Of the 76 catatonics, 38 or 50 per cent improved. The differential between catatonics and paranoids is in keeping with expectancy based on results obtained by other investigators [40, 33].

The sample contains 58 cases of affective disorders of various types. Of these, 41 or 70 per cent improved. The total number of cases is too small for any valid comparisons, but it is worth noting that the increased rate of improvement in contrast with the schizophrenic group is in accordance with expectancy both for the involutional and manic-depressive disorders [20, 17, 12].

Of the 147 nonpsychotic cases, 96 or 65 per cent improved. Ross [39], in a one-year follow-up of 1186 hospitalized neurotics, found 70 per cent improved. The several subtypes in this category contain too few cases for valid comparisons with other findings. It is worth noting, however, that the highest recovery rate is found for reactive depressions, which is in accordance with expectancy. [23].

These findings lend credence to the validity of our criteria, a crucial consideration in a study attempting to relate case-history variables to outcome of treatment. This correspondence between our findings and those of other investigators may be thought of as an estimate of the reliability of our criteria of improvement.

Criteria of improvement. A patient was considered improved if, one year after discharge from the WPIC, he was working at a level of efficiency roughly equivalent to that which obtained prior to illness. He was considered unimproved if, one year after discharge from the hospital, he was in another mental hospital, or, if not in a hospital, then incapacitated to the extent of being unable to be gainfully employed, or manifesting other behavior indicative of inability to cope with his environment in comparison with expectation of a person of like status. Only the dichotomy "improved"—"unimproved" was used. Borderline cases were judged, on the basis of available evidence, to fall in one category or the other. Most judgments, particularly for the improved, were made on the basis of observations made by psychiatrists or psychiatric social workers and recorded in the case records. The unimproved were judged, for the most part, on the basis of reports from other mental hospitals to which the patients had been transferred. This procedure was followed for 445 cases on whom follow-up material was available one year after discharge. Forty-one cases were judged by the following criteria: if the patient, on discharge from this hospital, was transferred to another hospital, unimproved with recommendation for commitment, and no follow-up information was available, the patient was judged unimproved; if the patient was discharged from this hospital, improved, and released to return to his previous or other similar environment, and no follow-up information was available, the patient was judged improved.

The variables. The variables used in the study represent a compromise between those variables deemed important from a consideration of evidence in the literature and the practical job of culling reasonably reliable information from the records. The variables are listed below. Definitions are given where indicated. In some cases only a brief, orienting definition is given; these are the variables which will be given later, fuller treatment.

1. Sex
2. Age at admission to WPIC
3. Marital status—three categories: single, married, divorced, or widowed
4. Religion—Catholic, Protestant, Jewish, or Greek Catholic
5. Nationality—of the patient or his parents: *American; North European* including France, Germany, British Isles, Low Countries, Scandinavia, Austria; *South European* including Italy, Balkan countries, Spain, and Portugal; *East European* including Russia, Poland, Czechoslovakia, Hungary, and Baltic States
6. Occupation—four categories: *professional,* including teachers, physicians, lawyers, engineers, scientists, artists, musicians; *semiprofessional,* including businessmen, salesmen, college students, bank tellers, dieticians, secretary (college graduates), etc.; *skilled,* including machinists, electricians, stationary engineers, cranemen, clerks, chauffeurs, stenographers, miners, high-school students; and *unskilled,* including day laborers, handymen, farmhands, domestic, unemployed, grade-school students
7. Economic level—four categories: comfortable, average, marginal, dependent
8. Body Type—only height and weight were available in the records. With these data only a crude estimate of body type could be attempted. Following Sheldon's suggestion [41] that height over cube root of weight would give an estimate of body type for the earlier years, this ratio was calculated for all patients under 30 years of age.
9. Education—calculated by grades
10. Intellectual status. IQ was not used. In an attempt to obtain a more valid measure of this variable, the clinician's estimate of the patient's level was taken rather than the IQ reported. Where the patient had not been seen by a psychologist, then the authors made a judgment based on the case-history material. Categories used were, above average, average, and below average.
11. Affective expression—a nine-point scale was constructed to rate this variable. Briefly, an attempt was made to rate the amount of affect displayed by the patient. This variable will receive detailed treatment in a later paper of the study [2].
12. Orientation—the psychiatrist's report as to whether or not the patient was oriented for time, place, or person was used to place the patient into one of two categories as follows: oriented, not oriented.
13. Direction of aggression — a nine-point scale was constructed to rate this variable. Briefly, an attempt was made to rate whether the direction of aggression displayed by the patient was towards himself or others. This variable will receive detailed treatment in a separate, later paper [10].
14. Diagnosis—the patient's final diagnosis, reached by agreement between the attending psychiatrist and the clinical director, or on the basis of a staff conference attended by several psychiatrists
15. Patient's age at onset of mental illness—the opinion of the attending psychiatrist and/or psychiatric social worker
16. Type of onset—a nine-point scale constructed to rate the continuum running from acute to insidious onset. This variable will receive detailed treatment in a separate paper [47].
17. Duration of the patient's illness—the opinion of the attending psychiatrist and/or psychiatric social worker and reported as such in the case record. This variable will be the subject of a later paper.
18. Total time of hospitalization—the total time, in months, spent by the patient in a mental hospital
19. Status on discharge from WPIC—the psychiatrist's opinion of the patient's condition immediately upon discharge. Categorized as improved or unimproved.
20. Precipitating stress—a nine-point scale constructed to rate the amount of environmental stress experienced by the patient immediately preceding the onset of illness. This variable will receive detailed treatment in a separate paper [6].
21. Amount of treatment received by the patient—three-point scale. Point *one*: routine hospital care, including psychiatric interviews, occupational therapy, and hydrotherapy. Point *two*: electroconvulsive therapy, insulin therapy, metrazol therapy, etc. Point *three*: somatic therapies plus prefrontal lobotomy.

This paper of the series will concern itself with reporting the relationship between each of these variables and the outcome of treatment, using appropriate statistical techniques. It will, in addition, report the interrelationships between those variables found to be significantly related to outcome of treatment. For this purpose a p value of .01 is considered indicative of a significant relationship. Later papers will report on single variables and their relationship to outcome of treatment when all the other variables found to be similarly related are held constant, insofar as possible. A final paper will present a discussion of the entire study, summary, and conclusions.

Results and Discussion

Each of the 21 variables was related to status one year after discharge from the hospital. Where the independent variable was categorized, chi square was used; where it was continuous, the biserial correlation coefficient was used. Table 2 shows the results of statistical calculations. Of the 21 variables, 10 seem to be unrelated to the criterion, with probability values of .02 or greater. These variables are sex, age, religion, nationality, occupation, economic level, body type, education, intellectual status, and age of onset. We shall not consider these variables further except to note that similar results are reported in the literature for each finding [43, 45, 42, 4, 16] except that for body type. For what it is worth the low relationship obtained tends to agree with Malamud and Render [28] and Kline and Tenny [25], who report a tendency for patients high in endomorphy to have a poor prognosis.

Table 2

Independent Variables Related to Status One Year after Discharge from WPIC

Variable	Statistic	p Value
1. Sex	chi square = 5.59	.02
2. Age	r_{bis} = .04	.05
3. Marital status	chi square = 12.87	.01
4. Religion	chi square = 7.48	.10
5. Nationality	chi square = .90	.90
6. Occupation	chi square = 5.12	.20
7. Economic level	chi square = 4.57	.30
8. Body type	r_{bis} = .08	.05
9. Education	r_{bis} = .10	.05
10. Intellectual status	chi square = 3.30	.20
11. Affective expression	r_{bis} = .32	.01
12. Orientation	chi square = 16.54	.001
13. Direction of aggression	r_{bis} = .67	.01
14. Diagnosis	chi square = 49.45	.01
15. Age of onset	r_{bis} = .09	.05
16. Type of onset	r_{bis} = .31	.01
17. Duration of illness	r_{bis} = .38	.01
18. Hospital stay	r_{bis} = .37	.01
19. Status on discharge	chi square = 244.65	.001
20. Precipitating stress	r_{bis} = .25	.01
21. Treatment	chi square = 16.42	.001

At least two impressive recent studies, however [29, 8], and a third older study [21] suggest that outcome of treatment may be related to intellectual status. We suggest the possibility that if obtained IQ's are used as an estimate of intellectual status, a relationship may be found between IQ scores and unfavorable outcome, because of the possible relationship between severity of illness and psychological deficit [19, 37, 35]. The discrepancy between the present findings and those reported elsewhere may be due to the fact that in the present study an attempt was made to assess intellectual status prior to the onset of illness.

Eleven variables were found to be related to status one year after discharge at acceptable levels of confidence. Table 4 shows the pertinent interrelationships between variables. Five of these, *affective expression, direction of aggression, type of onset, duration of illness, and precipitating stress,* will be the objects of detailed study and presented in separate papers. The other six will be dealt with here.

Marital status. The results obtained, expressed in percentages, are contained in Table 3. These results suggest that the divorced or separated and the single patients tend to have a less favorable prognosis than the married ones. These results accord with expectation based upon other reports in the literature [30, 16, 21, 28]. Wall [49] suggests that history of some attempt at sexual adaptation tends to indicate a good prognosis.

Table 3

Percentages of Patients Improved and Unimproved for Marital Status, Orientation, Diagnosis, Discharge Status, and Treatment

Criterion	N	Status 1 yr. after discharge Imp.	Unimp.
Marital Status			
Single	275	41%	59%
Divorced, Separated	41	39	61
Married	170	56	44
Orientation			
Oriented	413	50	50
Disoriented	73	27	73
Diagnosis			
Schizophrenia	264	34	66
Other psychoses	75	71	29
Nonpsychotic	147	65	35
Discharge status			
Improved	257	80	20
Unimproved	229	8	92
Treatment Rating			
1	147	47	53
2	300	50	50
3	39	16	84

Orientation. This variable seems to be significantly related to status one year after discharge. The disoriented patients, as defined, tend to have an unfavorable prognosis. Little relationship is found between orientation and outcome if the patient is oriented, as the tabulation in Table 3 indicates. A number of previous reports can be found which tend to support these findings [52, 51, 49, 27]. On the other hand, an even greater number of previous studies seem to argue against acceptance of the present findings [31, 22, 16, 13, 5, 4, 3]. The weight of the evidence in the literature seems to favor a relationship between good prognosis and confusion, clouding, perplexity, bizarreness of behavior, in fact, all the behaviors which might contribute to what, in the present study, has been categorized as disorientation. Orientation is, we feel, the kind of variable that will give results which are apt to fluctuate, depending on other characteristics of the population sample.

The prognostic value of disorientation probably depends upon how and in what context of other variables disorientation is displayed. Thus if delusions accompany disorientation, then there is evidence to indicate the type of delusions displayed influences prognosis [1]. If hallucinations contribute to disorientation, then there are indications that visual hallucinations are less ominous than auditory [51]. Table 4 indicates that orientation is significantly related to a number of other variables which are themselves significantly related to outcome. We might, for the sake of example, take the relationship found between direction of aggression and orientation. For those patients categorized disoriented, 70 per cent directed their aggression towards others, 23 per cent did not direct their aggression primarily in any one direction, and 7 per cent directed their aggression towards themselves. The implication here is clear that disorientation in our sample is related to behavior indicative of projection. In the same way it could be shown that disorientation is related to lack of precipitating stress, lack of affective expression, and the diagnosis of schizophrenia. We do not feel, in other words, that the fact that a patient is oriented or disoriented for time, place, and person at the time of the mental status is, by itself, a particularly significant datum upon which to base prognosis. It is difficult to see how a clear notion of the relationship between orientation

Table 4

Interrelationships between Significant Variables with Probability Values‡

	Aff. Exp.	Orient.	Dir. Agg.	Diag.	Onset	Dur. Ill.	Prec. Stress	Treat.
Affect. exp.	....	*61.2*	*.24**	*109.7*	.06* >.05	*26.9*	*.18**	*15.7*
Orient.	*61.2*	....	*12.8*	*55.9*	*10.6*	4.2 >.05	6.5 <.05	*25.4*
Dir. agg.	*.24**	*12.8*	....	*132.9*	*.20**	*24.3*	*.23**	*18.9*
Diagnosis	*109.66*	*55.9*	*132.9*	....	11.9 <.02	11.3 >.05	12.9 <.02	*57.6*
Type onset	.05* >.05	*10.6*	*.20**	*11.9*	....	*.33*	.11* <.05	*19.3*
Dur. ill.	*26.9*	4.2 >.05	*24.3*	11.26 >.05	*79.8*	....	*19.3*	*23.5*
Precip. stress	*.18**	6.5 <.05	*.23**	12.9 <.02	.11* <.05	*19.3*	....	.81 >.50
Treatment	*15.7*	*25.4*	*18.9*	*57.6*	*19.3*	*23.5*	.81 >.50	....
Marital status	*35.18*	4.21 <.20	9.41 <.05	5.84 <.10	6.10 <.05	4.25 <.20	*22.24*	1.44 <.50

*Correlation coefficients. All other values are chi squares.

‡Italicized values indicate significant relationships with probability values less than .01; other p values are shown.

and outcome can be obtained unless the variables related to it and outcome are controlled. By the same token, the significant relationship found between orientation and outcome points to the need to control for this variable in studying the relationship between other variables and outcome.

Diagnosis. The patient's diagnosis is significantly related to status one year after discharge. Table 3 summarizes the findings. These findings are in accordance with expectancy and corroborating studies have already been cited. Table 4 shows that diagnosis is also significantly related to affective expression, orientation, direction of aggression, and amount of treatment given. Of the three categories, schizophrenics tend to have less affective expression; the direction of their aggression is towards others, they tend to show more disorientation, and to be given more treatment. Each of these variables must be controlled in any careful study of the relationship between diagnostic category and outcome of treatment. It is our guess that if such controls are imposed, it is the relationship between diagnostic category and such variables as affect, direction of aggression, type of onset, etc., which gives diagnosis its prognostic validity. It is worth noting that, in this study, a nonsignificant relationship was found between diagnostic category and duration of illness.

Total hospitalization. This variable is significantly related to status one year after discharge. The longer the hospital stay the more tendency for the patient to remain unimproved one year after discharge. This variable is so highly related to duration of illness and the same variables which are related to duration of illness that we shall not consider it further, here, except to note that total hospitalization needs to be controlled when the relationship between other variables and outcome is studied.

Status on discharge. There is a consistent tendency for those who are discharged improved to remain improved for one year after discharge. The same tendency applies to those discharged unimproved. The prediction of status one year after discharge from status on discharge is quite good, as indicated by the summary of findings in Table 3. The implication, here, is that the results of this study might also apply fairly well to the immediate outcome of treatment, although with a lesser degree of accuracy.

Treatment. Total treatment, as crudely rated in this study, seemed to be significantly related to outcome. As the summary in Table 3 indicates, however, most of its significance derives from the great percentage of unimproved in the category which comprises the lobotomies. There seems to be little relationship between treatment ratings *one* and *two*, and outcome. A number of studies can be cited which corroborate this finding [14, 18, 15, 9, 5, 3], i.e., that there is little relationship between amount of somatic treatment and prognosis. A study by Danziger and Kindwall [7], however, suggests that the outcome of treatment in dementia praecox depends on the number of treatments either of insulin or electroconvulsive therapy, when duration of illness is held constant. Unfortunately, our treatment data are too crude to enable us to check on this very interesting finding with any degree of accuracy.

Table 4 shows that the treatment rating is significantly related to affective expression, orientation, direction of aggression, diagnosis, type of onset, and duration of illness. By and large, these relationships hold even when the lobotomies (treatment rating *three*) are excluded as Table 5, relating to treatment rating, indicates.

Table 5
Treatment Received by Patients with Various Diagnoses

		Treatment Rating	
		1	2
Diagnosis	*N*	routine hosp. care	physical therapy
Schizophrenia	232	19%	81%
Other psychoses	71	31%	69%
Nonpsychotic	144	57%	43%

It is evident, therefore, that if the relationship between other variables and outcome is so studied, the treatment rating needs to be held constant.

Summary and Conclusions

Using predetermined criteria, 486 case records were selected from the files of the Western Psychiatric Institute and Clinic. These records comprised all the cases meeting the criteria up to July, 1950. All were records of patients between the ages 15 and 55, without mental deficiency, without known damage to the cortex, and without record of previous attacks of mental illness. The patients' statuses one year after discharge from the hospital were determined and categorized improved or unimproved. The records were examined and 21 variables extracted which were, by appropriate statistical techniques, related to status one year after discharge. Of the 21 variables, 11 were found to be of prognostic significance at the 1 per cent level of confidence. These are marital status, affective expression, orientation, direction of aggression, diagnosis, type of onset, duration of illness, length of hospital stay, status on discharge, precipitating stress, and amount of treatment. The interrelations between these variables are presented and discussed.

The primary purpose of this study is to demonstrate methodology. Emphasis is given to the demonstrated need to control for other significant variables while the prognostic significance of a single variable is being studied. The present section of the study has been concerned with the presentation of significant variables and their interrelationships. Other following sections will present carefully controlled studies of well-defined, single variables and their relationship to status one year after discharge. A final section will summarize the findings and, by statistical techniques, suggest the most efficient combination of prognostic variables.

Received October 6, 1952.

References

1. Albee, G. W. The prognostic importance of delusions in schizophrenia. *J. abnorm. soc. Psychol.*, 1951, **46**, 208–212.
2. Bayard, Jean. Studies of prognostic criteria in the case histories of hospitalized mental patients. V. Affective Expression. Unpublished master's thesis, Univer. of Pittsburgh, 1952.
3. Bellak, L. *Dementia praecox.* New York: Grune & Stratton, 1948.
4. Chase, L. S., & Silverman, S. Criteria in schizophrenia: critical survey of literature. *Amer. J. Psychiat.*, 1941, **28**, 360–368.
5. Cheney, C. W., & Clow, H. E. Prognostic factors in insulin shock therapy. *Amer. J. Psychiat.*, 1941, **97**, 1029–1039.
6. Cole, Mary E. Studies of prognostic criteria in the case histories of hospitalized mental patients. III. Precipitating stress. Unpublished doctor's dissertation, Univer. of Pittsburgh, 1952.
7. Danziger, L., & Kindwall, J. A. Prediction of immediate outcome of shock therapy in dementia praecox. *Dis. nerv. Syst.*, 1946, **7**, 299–303.
8. Dewan, J. G. Intelligence and emotional stability. *Amer. J. Psychiat.*, 1948, **104**, 548–554.
9. Dukor, B. Prognosis. *Schweiz. med. Wchnschr.*, 1939, **69**, 45–92.
10. Feldman, Dorothy. Studies of prognostic criteria in the case histories of hospitalized mental patients. II. Direction of aggression. Unpublished doctor's dissertation. Univer. of Pittsburgh, 1952.
11. Fishbein, I. L. Involutional melancholia and convulsive therapy. *Amer. J. Psychiat.*, 1949, **106**, 128–135.
12. Gildea, E. F., & Man, E. B. Methods of estimating capacity for recovery in patients with manic-depressive and schizophrenia psychoses. *Amer. J. Psychiat.*, 1943, **99**, 496–506.
13. Gold, L., & Chiarello, C. J. Prognostic value of clinical findings in cases treated with electric shock. *J. nerv. ment. Dis.*, 1944, **100**, 577–583.
14. Gottlieb, J. S., & Huston, P. E. Treatment of schizophrenia: followup results in cases of insulin shock therapy and in control cases. *Arch. Neurol. Psychiat.*, 1943, **49**, 266–271.
15. Gralnick, A. Seven-year survey of insulin treatment in schizophrenia. *Amer. J. Psychiat.*, 1945, **101**, 449–452.
16. Herzberg, F. I. Prognostic variaties for electro-shock therapy. Unpublished doctor's thesis, Univer. of Pittsburgh, 1950.
17. Hohman, L. B. A review of 144 cases of affective disorders after seven years. *Amer. J. Psychiat.*, 1937, **94**, 303–308.
18. Horwitz, W. A., & Kleimann, C. Survey of cases of dementia praecox discharged from Psychiatric Institute and Clinic Hospital. *Psychiat. Quart.*, 1936, **10**, 72–86.
19. Hunt, J. McV., & Cofer, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944. Pp. 971–1032.
20. Hunt, R. C., & Appel, K. E. Prognosis in psychoses lying midway between schizophrenia and manic-depressive psychoses. *Amer. J. Psychiat.*, 1936, **93**, 313–339.
21. Jacob, J. S. Prediction of outcome-on-furlough of dementia praecox patients. *Genet. Psychol. Monogr.*, 1940, **22**, 425–453.
22. Kant, O. The evaluation of prognostic criteria in schizophrenia. *J. nerv. ment. Dis.*, 1944, **100**, 598–605.
23. Karagulla, S. Evaluation of electric convulsion therapy as compared with conservative methods of treatment in depressive states. *J. ment. Sci.*, 1950, **96**, 1060–1091.

24. Kitching, E. H. The prognosis of mental disorders. *Clin. J.*, 1948, **77**, 41–51.

25. Kline, N. S., & Tenney, A. M. Constitutional factors in the prognosis of schizophrenia. *Amer. J. Psychiat.*, 1950, **107**, 434–441.

26. Lewis, A. J. Prognosis in schizophrenia. *Lancet*, 1935, **1**, 339–341.

27. Lewis, N. D. C., & Hubbard, L. D. The mechanisms and prognostic aspects of the manic-depressive–schizophrenic combinations. Manic-Depressive psychosis. *Proc. Ass. Res. nerv. ment. Dis.*, 1931, **11**, 539–608.

28. Malamud, W., & Render, N. Course of prognosis in schizophrenia. *Amer. J. Psychiat.*, 1939, **95**, 1039–1057.

29. Malamud, W., Sands, S. L., Malamud, Irene T., & Powers, P. J. P. The involutional psychoses: a socio-psychiatric follow-up study. *Amer. J. Psychiat.*, 1949, **105**, 567–572.

30. Malzberg, B. Marital status in relation to the prevalence of mental disease. *Psychiat. Quart.*, 1936, **10**, 243–261.

31. Meduna, L. J. *Oneirophrenia.* Urbana: Univer. of Illinois Press, 1950.

32. Neumann, E., & Finkenbrink, F. (Statistical studies in spontaneous remissions in schizophrenia.) *Allg. Zchr. f. Psychiat.*, 1939, **111**, 17–46.

33. Osborne, R. L. Prognosis in schizophrenia. *J. Amer. med. Ass.*, 1940, **114**, 846–847.

34. Palmer, D. M., Riepenhoff, J. P., & Hanahan, P. W. Insulin shock therapy, a statistical survey of 393 cases. *Amer. J. Psychiat.*, 1950, **106**, 918–926.

35. Pascal, G. R., & Zeaman, J. B. Measurement of some effects of electro-convulsive therapy on the individual patient. *J. abnorm. soc. Psychol.*, 1951, **46**, 104–115.

36. Paster, S., & Hotzman, S. C. A study on one thousand psychotic veterans treated with insulin and electric shock. *Amer. J. Psychiat.*, 1949, **105**, 811–814.

37. Rapaport, D. *Diagnostic psychological testing.* Chicago: Year Book Publishers, 1945.

38. Rennie, T. A. C. Prognosis in manic-depressive and schizophrenic conditions following shock treatment. *Psychiat. Quart.*, 1943, **17**, 642–654.

39. Ross, T. A. *An enquiry into prognosis in the neuroses.* Cambridge: University Press, 1936.

40. Rupp, C., & Fletcher, E. K. Five to ten years follow-up study of 641 schizophrenic cases. *Amer. J. Psychiat.*, 1940, **96**, 877–888.

41. Sheldon, W. H., Stevens, S. S., & Tucker, W. B. *The varieties of human physique.* New York: Harper, 1940.

42. Stallworthy, K. R. Prognosis in schizophrenia. *New Zealand Med. J.*, 1944, **43**, 230–232.

43. Strecker, E. A., Appel, K. E., Eyman, E. V., Farr, C. B., LaMar, N. C., Palmer, H. D., & Smith, L. H. The prognosis in manic-depressive psychosis. Manic-depressive psychosis. *Proc. Ass. Res. nerv. ment. Dis.*, 1931, **11**, 471–538.

44. Strecker, E. A., & Willey, G. F. An analysis of reversible "dementia praecox" reactions. *Amer. J. Psychiat.*, 1924, **80**, 592–677.

45. Strecker, E. A., & Willey, G. F. Prognosis in schizophrenia. Schizophrenia. *Proc. Ass. Res. nerv. ment. Dis.*, 1928, **5**, 403–431.

46. Sullivan, H. S. The relation of onset to outcome in schizophrenia. Schizophrenia. *Proc. Ass. Res. nerv. ment. Dis.*, 1931, **10**, 111–118.

47. Swensen, C. H. Studies of prognostic criteria in the case histories of hospitalized mental patients. IV. Type of onset. Unpublished doctor's dissertation, Univer. of Pittsburgh.

48. Taylor, S. D., & Von Salzen, C. F. Prognosis in dementia praecox. *Psychiat. Quart.*, 1938, **12**, 576–582.

49. Wall, C. Some prognostic criteria for response of schizophrenic patients to insulin treatment. *Amer. J. Psychiat.*, 1941, **97**, 1397–1402.

50. Wilcox, P. H. Shock therapy. In E. A. Spiegel (Ed.), *Progress in neurology and psychiatry.* New York: Grune & Stratton, 1949. Pp. 499–528.

51. Williams, R. R., & Potter, H. W. The significance of certain symptoms in the prognosis of dementia praecox. *State Quart.*, 1921, **6**, 361–380.

52. Wittman, Phyllis. A scale for measuring prognosis in schizophrenic patients. *Elgin St. Hosp. Papers.*, 1941, **4**, 20–33.

53. Wittman, Phyllis, & Steinberg, D. L. Follow-up on objective evaluation of prognosis in dementia praecox and manic-depressive psychoses. *Elgin St. Hosp. Papers.*, 1944, **5**, 216–227.

Journal of Consulting Psychology
Vol. 17, No. 3, 1953

An Evaluation of Dymond's Test of Insight and Empathy

Henry Clay Lindgren
San Francisco State College

and Jacqueline Robinson [1]
San Francisco, California

Most persons who have to do with selection, training, or supervision agree that the crucial factors which differentiate between good and poor workers have to do with personality rather than skill and information. This is particularly true in the case of workers in such fields as teaching, social welfare, nursing, and counseling and guidance. However, we have had relatively little success so far in devising methods and devices which can measure the dimensions of personality that are essential to professional success with any great degree of validity and reliability.

This study is the report of an attempt to develop a method of measuring insight and empathy—two qualities or abilities essential to the successful conduct of any profession concerned primarily with interpersonal relations.

The test or method which appeared promising is one devised by Rosalind Dymond [3, 4]. Dymond employed a cross-questionnaire type of test administered to a group of 80 people, subdivided into smaller groups of 4 each. After a brief acquaintanceship period, each member of the subgroup was asked to compare himself on a five-point scale with each other member of the group with regard to six different qualities: superiority, self-assurance, leadership, friendliness, sympathy, and tension. He was asked to repeat these six comparisons in a slightly different way in each of the four sections of each questionnaire. In the first section, he was asked to compare himself to another person in the subgroup, whom we shall call the *referent*; in the second section, he was asked to compare the referent to himself; in the third section he was to guess what the referent would say ("What will he say he is compared to you?"); and in the fourth section he was asked to guess the reverse of this question ("What will he say you are, compared to him?"). After each person in the subgroups had rated or compared himself with each other person in his subgroup, the subgroups were rearranged and re-formed in such a way that the new subgroups consisted of individuals who had not been used as co-referents. The process of comparison and rating was repeated. In all, the process was performed three times. In this way, each subject became a member of three different groups and interacted with nine different referents.

The "empathy" score for each subject was obtained by comparing his predictions of what referents would say (Sections III and IV of the questionnaire) with the actual statements of the referents (Sections I and II of the referents' questionnaires). The empathy score was the total number of scale-points the predictions deviated from the actual responses. Thus, the higher the score (i.e., the greater the error), the less the "empathy."

If the empathy is used in the sense employed by Dymond, i.e., the " 'faculty' of being able to see things from the other person's point of view," [4, p. 344] this method possesses a certain face validity, in that the individual is asked to empathize, to predict the thoughts or feelings of another, and a check is made to see whether he has empathized correctly.

The "insight" score was obtained by comparing a subject's judgments about himself (Section I) to the judgments which others made

---

[1] The authors wish to acknowledge the help of graduate students at San Francisco State College who aided in assembling the data on which this study is based: Charles Smith, Jack Nakashima, Donald Kase, Donald Crosby, and Richard Gilberg.

about him (Section II). The concept of insight employed by Dymond was Allport's criterion: the relation of what an individual thinks he has to what others think he has [1]. Although many would undoubtedly object to the narrowness of this definition of insight, it is to be granted that Dymond's method of measurement of this factor also has face validity, provided this definition is kept in mind.

A Revision of the Dymond Test

The first step to be undertaken in determining whether this test could be used as a selection device was to try it out in a large-scale administration. Dymond used a complete questionnaire of some nine pages for each individual and each administration. This meant nine questionnaires or 81 pages for each individual. It was evident that the bulk of paper would be large and the scoring cumbersome if her method were employed without change. Her method was therefore simplified by constructing an answer sheet containing the statements as originally made on the questionnaire, but with the five-point scale rendered as "*A B C D E.*" Thus each individual used the basic questionnaire much like a test booklet and recorded his replies on the answer sheet.

This format was tried with two groups of freshman psychology students with a total enrollment of 60. The mean and standard deviation of 71.3 and 15.9 of the empathy scores compare favorably with the mean of 73.2 and the standard deviation of 15.8 reported by Dymond.

However, students taking the test were quite critical of the items relating to "superiority," which they stated had no meaning for them, unless it was one of snobbery. Inasmuch as the "superiority" statement was the first item in each section, it was felt by the experimenter that it might possibly have an adverse effect on the cooperation of subjects. Consequently, the questionnaire and answer sheet were revised so that the item relating to "superiority" was changed to one of "intelligence" and assigned a medial position in the four sections.

The second revised form of the questionnaire was administered to two groups of freshmen and one group of upper-division students, 125 in all. The mean for this group was 71.23, and the standard deviation was 17.55. Here again, figures compare very favorably with the scores originally obtained by Dymond.

Reliability and Validity

A number of checks were made on the validity and reliability of the test. Dymond had reported a correlation of .63 between the first half and the second half of the empathy scores on each sheet, and a correlation of .82 between scores on items 1, 3, and 5, and scores on items 2, 4, and 6 [4]. She also reported a test-retest reliability of .60 for an earlier form of her test [3].

The reliability of the present test was computed by the split-half method, i.e., by dividing each person's answer sheets into two groups on an odd-even basis and correlating the total scores for each of the two groups. The nine sheets made two groups of five each, the even-numbered sheets being supplemented by one sheet selected at random from the "odd" group. The correlations obtained were .69 for the empathy portion of test and .73 for the insight portion ($N = 87$).

Dymond had also reported that the Wechsler full scale IQ of "good empathizers" was 132.1 and that of "poor empathizers" was 126.4, a difference which she described as not significant [4]. Inasmuch as she did not report correlations between Wechsler and empathy scores, it was difficult to make comparisons. Nevertheless, correlations were computed for 75 students between scores received on the revised form of the Dymond test and the American Council on Education's Psychological Test for College Freshmen, with the following results: correlation between "empathy" and the ACE (*L*), .14; between "empathy" and ACE (*Q*), .17; between "empathy" and ACE (Total), .14; and between "insight" and ACE (Total), .02. These results confirm Dymond's interpretations.

An investigation was also made of the relationship between the individual's rating of his own empathy and the score he received on the empathy portion of the revised Dymond test. The correlation obtained for 48 individuals was −.04.

Another investigation was made of 45 students who had taken the Minnesota Multiphasic Personality Inventory as well as the revised form of the Dymond test. Although the relationship between empathy scores and scores received on the various MMPI scales was not thoroughly investigated, it was noted that a group consisting of the "poorest empathizers" received consistently higher (i.e., more "maladjusted") scores on the MMPI, than did a group consisting of the "best empathizers." These results are presented in Table 1. They tend to confirm the general nature of Dymond's findings that persons with "poor" empathy are not as well adjusted, according to personality tests, as are persons with "good" empathy [4].

Table 1

MMPI Scores* of "Good" and "Poor" Empathizers

MMPI Scales	"Good" Empathizers	"Poor" Empathizers
L	45	48
F	52	65
K	53	48
Hs	46	51
D	56	60
Hy	52	56
Pd	57	67
Pa	62	74
Pt	54	64
Sc	57	67
Ma	60	66
Number†	7	5

* Mean standard scores, including "K-correction" where appropriate.

† Out of a total of 45 freshmen.

The Effect of Cultural Stereotypes

In scoring the revised form of the Dymond test, the authors of this study became aware of the fact that responses to each item did not distribute themselves evenly with regard to the five points on each scale. There was a definite tendency for one of the positions in the scale to be preferred with regard to each item. For example, with regard to the item on friendliness ("How friendly or unfriendly do you think you are to him or her?"), there was a decided tendency to check the fourth item ("fairly friendly"). This observation gave rise to the hypothesis that perhaps respondents were actually not empathizing at all, but were, in effect, responding to a cultural norm or stereotype which says that people ought to maintain attitudes which are "fairly friendly." To carry this hypothesis further, it is implied that individuals in our middle-class culture would be under some pressure either to behave in ways which are "fairly friendly" or to *believe* that they are "fairly friendly," i.e., to conceive of themselves as "fairly friendly" even if this were not actually characteristic of their behavior. "Friendliness" occupies a fairly central position in the values of our culture; we have a high stake in conforming to the cultural norms; and each of the personal qualities which form the bases of the items in the questionnaire likewise occupies a central position with regard to the norms and values of our culture.

When this hypothesis was checked with regard to responses indicated on the answer sheets, a rather decided pattern emerged. According to the patterning of the responses on the first section, we are likely to think of ourselves as "fairly self-assured" in our relations with others; we are "about half and half leaders and followers"; we are "fairly friendly"; we are "neither more nor less intelligent"; we are "fairly sympathetic"; we are "fairly relaxed." Furthermore, a similar sort of patterning persisted throughout the remaining three sections of the questionnaire.

On the basis of this analysis, a male key and a female key were constructed based on the responses of 100 students, equally divided as to sex.[2] Tests were then rescored in two different ways. They were scored by the keys: *1.* According to the original scoring patterns, i.e., an "empathy" score was obtained by measuring the amount of deviation between Sections III and IV of the subject's questionnaires and Sections I and II of the key, and "insight" was measured by comparing Section I of the subject's questionnaires with Section II of the key. *2.* The second method of scoring was to measure the deviation between each section of the subject's questionnaires and the corresponding section of the key. This method of scoring produced a score which represented the tendency of an individual to conform to the norm in his responses.

---

[2] To save printing costs, the questionnaire, answer sheet, and keys have been deposited with the American Documentation Institute. Order Document No. 3916 from the American Documentation Institute, Auxiliary Publications, Photoduplication Service, c/o Library of Congress, Washington 25, D. C., remitting $1.25 for microfilm (images 1 inch high on standard 35 mm. motion picture film) or $1.25 for photostat readable without optical aid.

Using the first method of scoring, the correlation between the original empathy scores and the "normative" empathy scores was .74. The correlation for insight was .60.

Using the second method the correlation between the original empathy scores and the total normative score was .56. The correlation with insight scores was .51.

Discussion

The results of this investigation indicate that the Dymond approach to the measurement of insight and empathy can be adapted to large-scale administration, but the reliability is somewhat low for the test to be considered useful as a predictive device. However, reliability could probably be increased to a satisfactory level by increasing the number of individuals with whom the subject rates himself. On the other hand, such a procedure introduces further complications. The present manner of administration requires somewhat more than a two-hour period, or three fifty-minute class periods. Furthermore, it is difficult to maintain the interest of subjects even for the three administrations required. On the other hand, if the test were actually used as a selection instrument, presumably the ego-involvement of participants would be sufficient to maintain interest even if the number of administrations were doubled.

The validity of the test is in graver question. If it measures insight according to the definition selected by Dymond, it probably covers too narrow an aspect of what psychologists mean by insight. What we need here is a term which means "the tendency to see oneself as seen by others," since this is what the test measures.

Further data are needed as to the extent to which the test measures empathy. Dymond's research as well as that of the present study indicates that empathy as measured by this test is probably related to personal adjustment as measured by a variety of measures. However, what is needed is an evaluation based on more acceptable criteria of empathy. The content-free interview method of measuring empathy developed by Claire Wright Thompson and Katherine Bradway suggests itself as a possible criterion [2].

The gravest question with regard to validity grows from the relationship demonstrated in this study between the "normative" scores and scores produced by the method developed by Dymond. This relationship comes close to the reliability of the test. In other words, the test correlates almost as well with the cultural norm as it does with itself. This raises the question of what an individual does who compares himself with others by the Dymond method. Is he responding to his awareness of others, or is he comparing his self concept with some cultural norm, or is he reacting to some stereotyped concept of how people ought to feel and act? Or is he reacting defensively because the nature of the questions arouses anxiety?

Furthermore, one should not overlook the possibility suggested by Hastorf and Bender that "part of the successful prediction of another person's responses may be due to projection rather than empathy. . . ." [5, p. 576]. This hypothesis could account for the high correlation between empathy scores and normative scores reported in the present study. It may be, therefore, that persons show up as "good empathizers" on this test if they are conventional folk who conform rather closely to cultural norms and standards and who tend to perceive the behavior and attitudes of others in the light of these same norms. In other words, conventional people get good scores on empathy tests because most of their partners (or referents) in the test are also conventional. Hence the prediction of conventional attitudes coincides with the existence of conventional attitudes. This would also mean that unconventional people would tend to get poor empathy scores (because they would project deviating attitudes). Thus when Dymond states that "on the whole it is easier to predict the responses of a person who is highly empathic" [4, p. 345] as measured by this test, she may be reporting the discovery that the more conventional an individual is, the easier it is to predict his responses.

It is quite possible that all these factors operate. In any case, further research with instruments of this type will have to take into consideration the effect of norms and stereotypes on the results of the test.[3] Perhaps the empathy of an individual can be measured by scoring his questionnaires by both the normative key and by the Dymond method. Using this approach, an individual who has a "good" (i.e., low) score on the empathy portion of the test, but who has a high score (showing marked deviation) from the normative key would be a person with much empathy, little influenced by the cultural norm. However, it would be difficult to demonstrate whether a person who receives a low score on each of these scorings actually lacks empathy. Perhaps what is needed,rather, is a test constructed on the same model, consisting of items which are less ego-involved, which do not evoke defensiveness or references to cultural norms. However, the construction of such a test would not be easy, as any item involving interpersonal relationships would be likely to incur some anxiety and hence some defensiveness. Nevertheless, such a test would have great value as a selection instrument for occupations in which empathy is important. Thus it is hoped that additional research will produce an instrument which is valid and reliable for this purpose.

Summary

A minor revision of the method used by Rosalind Dymond in the measurement of insight and empathy was used in an attempt to produce an instrument which could be used for purposes of selection. Dymond's findings were confirmed that the relationship between the factors measured by her test and scores received on intellective measures was slight. Also confirmed were her findings that persons of high empathy have "better" scores on personality tests than do persons of low empathy. However, the correlation between repeated administrations of the revised questionnaire indicated reliability of the method is too low for predictive purposes, although it probably would be improved if the total number of administrations were increased.

However, the validity of the method is questioned in view of the high degree of relationship between scores received on the revised test and those produced by scoring with a normative key. The correlation between these scores is almost as high as the reliability of the test. This raises the question of whether the test measures the tendency of individuals to respond to an interpersonal situation in terms of cultural norms rather than empathic promptings. It was suggested that both factors may operate. Yet until they can be measured separately, or until a form of the test is developed which does not evoke reference to such norms, the use of the present revision as a predictive measure of insight or empathy appears inadvisable.

Received October 6, 1952.

References

1. Allport, G. W. *Personality: A psychological interpretation.* Boston: Houghton Mifflin, 1937.
2. Bradway, Katherine, & Thompson, Claire Wright. A content-free method of training interviewers. *J. consult. Psychol.*, 1950, **14**, 321–323.
3. Dymond, Rosalind F. A scale for the measurement of empathic ability. *J. consult. Psychol.*, 1949, **13**, 127–133.
4. Dymond, Rosalind F. Personality and empathy. *J. consult. Psychol.*, 1950, **14**, 343–350.
5. Hastorf, A. H., & Bender, I. E. A caution respecting the measurement of empathic ability. *J. abnorm. soc. Psychol.*, 1952, **47**, 574–576.

---

[3] In a letter to the senior author, Dymond states that she has developed a partial-correlation method which partials out the stereotype from predictions, thus leaving a residual which may have more validity as a measure of empathy.

Journal of Consulting Psychology
Vol. 17, No. 3, 1953

The Present Status of Research on Nondirective Play Therapy

Dell Lebo
Florida State University

To many persons nondirective play therapy has seemed easy to learn, pleasant to undertake, and gratifying in results. Much of the attention attracted to play therapy has resulted from its apparent ease as well as from the concept that play is the natural medium of expression of the child. Play has come to be recognized as the most satisfactory way of understanding the nonverbal child. Because of the current widespread interest in child psychology, child training, education, and mental hygiene, recent years have seen a great deal of concomitant activity in nondirective play therapy. Much of the activity has resulted in emotional articles lauding nondirective play therapy. These articles generally explain the efficacy of nondirective play therapy on the basis of philosophical constructs arising from the growth principle developed from nondirective counseling with adults.

Research in nondirective therapy with adults is sound and extensive. Research in nondirective play therapy with children is still meager, unsound, and frequently of a cheerful, persuasive nature. It has seemed to the present writer that such articles could be more correctly classified as propaganda than as research.

The present paper is an attempt to review current research in nondirective play therapy. With one exception, studies involving more than one child have been reported here. The report giving only a single case history has been avoided. Also, only studies concerned with nondirective play therapy are reported. There is no doubt that play therapy has a considerable history. It may be said to extend back to Rousseau [23] who studied the play of the child to understand his psychology. However, nondirective play therapy developed from the work of Carl Rogers [21, 22] and his associates. One of them, Virginia Axline [2], was the first successfully to apply nondirective methods to play therapy with children. Axline's book, while widely read, is more suggestive than it is factual. It seems to have been the forerunner of much of the persuasive material to be found in studies of nondirective play therapy.

A recent example of the continuing propagandistic tendencies in nondirective play-therapy research papers is to be seen in an article titled: "An Experiment in Play Therapy" [11]. The stated purpose was to help children who seemed unable to adjust to the school situation. The children were selected on the basis of Rorschach tests, their teacher's impression, and the therapist's observations. Two groups of five children each were selected for nondirective play therapy. The majority of children selected had problems of sibling rivalry. The groups met once a week for play therapy. Much verbatim material from the therapy notes was presented to enable the reader to partake of the emotional flavor of the situation.

The shortcomings of the "experiment" are serious. First, it lacks a clear hypothesis and a control group. Second, the results seem to indicate a lack of rigorous method. For it is the conclusion of the research that "through group work they [the children] learned that they were not alone in having 'bad feelings.' Other children had them too" [11, p. 180]. Further that ". . . their natural healthy drive toward maturity, which had been retarded, could once more assert itself" [11, p. 180]. Such conclusions would seem to savor more of a desire to support nondirective play therapy than they do of experimental procedure.

What takes place in nondirective play therapy? With the philosophy of the love of children, and the idyllic purposefulness of many of the typical articles stripped off, nondirective play therapy seems to be left a rather thin framework. A determination of the process of play therapy, as contrasted with the results of play therapy, has been subject matter for only three known research studies.

The first such study was the work of Landisberg and Snyder [18]. They attempted to analyze what actually took place in client-centered play therapy by an objective approach. Their procedure was to study the protocols of three successful and one incomplete case. Each statement made by the counselor was categorized as to its content. Statements made by the children were categorized as to content, emotion expressed, and activity. Although the children ranged in age from five to six years, the categories used had been developed for employment with adult cases.

They reported finding an increase in the child's physical activity during the last three-fifths of therapy. The children were found to have released much feeling during therapy. About 50 per cent of their actions and statements during the first two-fifths of treatment were devoted to emotional release. This percentage rose to 70 for the last three-fifths of the process. It was noticed that negative feelings particularly increased in frequency. The major part of the children's feelings were directed towards others and not to themselves or to the counselor. No insightful statements were made by the children whose records were studied.

Finke [15] did not use adult categories in analyzing children's nondirective play therapy protocols, but derived her categories from an analysis of children's statements. Expressions of feeling were emphasized as it was believed such expression would mirror the child's changing emotional reactions resulting from the play therapy.

She selected complete protocols from six play therapists concerning six different children referred for behavior problems. The children ranged in age from five to eleven years. The possibility of bias resulting from one person's categorizing all the cases was avoided by having five students recategorize one or more interviews chosen at random. Their results corresponded adequately with the original categorization.

It was found that different children, undergoing therapy with different therapists, showed similar trends which tended to divide play therapy into three stages:

1. Child is either reticent or extremely talkative. He explores the playroom. If he is to show aggression at any time during therapy, a great deal of it will be exhibited in this stage.
2. If aggression has been shown, it is now lessened. This child tests the limitations of the playroom. Imaginative play is frequently indulged in here.
3. Most of the child's efforts are now expended into attempted relationship with counselor. The child tries to draw the therapist into his games and play.

Like Landisberg and Snyder [18], Finke [15] found no trends for positive statements. Unlike them, she found no trends for negative statements. The verbal characteristics of adult counseling sessions did not appear. Finke concluded that nondirective play therapy had its own characteristic pattern which was repeated in case after case.

Both studies indicated that children's attitudes changed during therapy and that the changes could be quantitatively reported. There seem to be serious limitations in both studies. Landisberg and Snyder used adult categories. Finke found fault with this. She indicated that differences in the age and sophistication of the adults and the children would affect the degree and type of verbalization made. Consequently, she felt it was not justifiable to evaluate children's comments on the basis of categories derived from adults.

Finke seemed to have failed to recognize the possibility that just as the wide age discrepancies between adults and children might influence the character of their verbalizations, so might children's categories vary significantly from one level of maturity to another.

The present writer [19] undertook a study of the possible relationship between chronological age and the types of statements made by children in play therapy. He used Finke's [15] categories.

Twenty children were given three play-therapy sessions by the same therapist in the same playroom. The children were reasonably equated for intelligence and social adjustment. Five age stages were represented with two boys and two girls in each stage. Children were selected who were 4, 6, 8, 10, and 12 years of age.

Fifteen pages of verbatim style notes were selected by a table of random numbers from the 166 pages of protocol. These 15 pages were then categorized by three experienced play therapists. Their percentages of agreement were adequately similar to one another. All of the protocols were then analyzed by the writer.

It was found that maturation, as represented by chronological age, did seem to account for some definite trends in the types of statements made by children in the play-therapy situation.

As the children became older, they told the therapist fewer of their decisions. They spent less time in exploring the limitations. They made fewer attempts to draw the therapist into their play and they expressed more of their likes and dislikes.

The three studies, while not strictly comparable, would seem to indicate that nondirective play therapy is an objectively measurable process; that children's emotional expressions are altered in a discernible manner; and that maturation appears to be related to the type of expression of therapeutic change. Beyond such statements the studies substantiate few of the philosophical aspects of play therapy.

The Successfulness of Nondirective Play Therapy

The outcome of play therapy in various types of cases. Nondirective play therapy has been used in the study and treatment of such seemingly diverse problems as allergy, mental deficiency, personality problems, physically handicapped children, race conflicts, and reading difficulties. From the published reports one receives the impression that it has usually been either successful or incomplete. The children are seemingly relieved of their presenting symptoms or the therapy is unavoidably interrupted in a promising but unfinished phase.

Nondirective play therapy in the treatment of allergy. Miller and Baruch [20] following a successful preliminary psychotherapeutic treatment of allergy [8] undertook to treat six children under eleven years of age by play therapy. All their subjects had classical allergic symptoms confirmed by positive skin reactions to various allegens. Prior to nondirective play therapy all the subjects had been unsuccessfully treated medically.

They cite as a representative case a five-year-old asthmatic boy who used attacks of asthma to gain contact with his mother. Whenever she left him, the asthma would express his hostile feelings. His asthmatic attacks cleared after five months of play therapy. Unfortunately, for purpose of this investigation, 16 allergic adults were included in the results. As a result, it can only be said that of the 22 patients (including the six children), 21 showed improvement while one was unchanged.

Nondirective play therapy in the study of mental deficiency. Exploratory material suggestive of the emotional factors in mental deficiency is presented by Axline [5]. In a report of an examination of selected play-therapy protocols, evidence is offered which indicates marked improvement in some IQ scores after completing play therapy.

The verbatim stenographic reports of 15 six-and-seven-year-old children referred for behavior problems were studied. Each child had been seen individually by the same therapist for 8 to 20 contacts. The reports were selected and analyzed at some time after therapy on the basis of Stanford-Binet IQ ratings and the age of the children. The records were then grouped as follows:

1. Children who showed no appreciable change in IQ scores after therapy. Pre- and posttesting indicated low intelligence.
2. Children who showed a gain in IQ scores after therapy. Pretests were low; posttests were indicative of normal intelligence.
3. Children with average intelligence both before and after play therapy. Children from this group had play therapy in a children's home.

In every case of those children whose IQ stayed low, their mother had indicated shame, disapproval, and rejection. It was felt that the children's difficulty lay in their daily lives. "They were not able to communicate clearly to others the things that were uppermost in their lives" [5, p. 528]. The therapist felt that each of the cases in the first category was incomplete. However, it was impossible to finish therapy.

In the case material presented, it is evident that both the children whose IQ's did not improve and those whose IQ's became normal initiated play activity. Both groups freely expressed negative feelings and destructive play which was followed by outgoing and more positive behavior. The only difference would seem to be that the children whose IQ's were raised had completed their therapy.

The same behavior was shown by the children of average intelligence whose IQ's did not change. This group was included to indicate that mental deficiency was not the cause of behavior problems for these children.

Axline [5] did not claim that nondirective play therapy raised the IQ of the children of group two. She explained the increase in IQ scores by saying the child was freed from emotional constraint and could thus more adequately express his true capacities.

Nondirective play therapy in the treatment of personality disorders. While most of the work in play therapy has been done in the area of personality, there is a dearth of research material. This can be explained by the client-centered philosophy from which nondirective play therapy sprung. Play therapy is oriented around the needs of the client and not around the demands of research. For this reason there are "cases" offered to prove that play therapy works, but there are still few research studies undertaken to see how well it works. Bloomberg's [11] work, already discussed, represents the research aspects of proselytizing for play therapy. Instead of presenting one case, she presents ten. But she presents these ten cases in a manner in keeping with the philosophy of nondirective play therapy. That is to say, the experiment was not designed to stimulate research, it was designed to help children. Consequently, her material merely indicates that play therapy works — a fact which no one disputes.

An experiment that deserves to be a model for future play therapy work is available in this area. Fleming and Snyder [16] endeavored to determine if measurable changes in social and personal adjustment resulted from nondirective play therapy.

They had three simple personality tests administered to 46 children. Seven children who ranged in age from eight to eleven years were selected for play therapy on the basis of poor results in these tests. After a lapse of 12 weeks, 30 of the 46 children were available for retesting.

Fleming and Snyder [16] found the three girls had improved their adjustment with a greater amount of positive feelings. The least amount of improvement for the girls was in the social area. Save for one individual who fared worse, the four boys made no significant changes. The control group posttest score was the same as their pretest score.

From an analysis of the group scores as well as from individual data, they concluded the greatest change for the subjects was in personal feelings toward the self and in daydreaming. Hence, the theory is offered that personal changes in adjustment must precede social change. The therapy experience had created more positive feeling among the subjects but it did not cause the control group to like them any better.

Since this was the first study of its kind, it was to be expected that certain of their findings should contradict some of Axline's [2] early observations. Although Fleming and Snyder did not indicate it, the following contradictions of Axline's observations were suggested:

1. The therapist's sex was an important factor in establishing rapport. They found ten-year-old boys would not respond well to a female therapist. Axline had said, "Nor does the sex of the therapist seem to be important [for successful play therapy]" [2, p. 65].
2. A housemother who was not given therapy prevented successful therapy with the boys' group. Previously, Axline had stressed that, "It is not necessary for the adults to be helped in order to insure successful play-therapy results" [2, p. 68].
3. The best therapeutic results seem to be achieved when the children in the group have the same degree of maladjustment. Axline had said, "Experiments in groupings indicate that there are no . . . rules to govern them: Successful groupings have included both sexes, siblings, and wide age ranges" [3, Pp. 269–270]. In another place she noted, "A handicapped child can be treated in a group with normal children" [3, p. 27].

Axline's statements would seem to warrant additional experimental investigation.

Nondirective play therapy in the treatment of children with physical handicaps. Axline [2] included cases of handicapped children in her pioneer work. Cowen and Cruickshank [12, 13] undertook the only other known research study to supplement her reports. They set themselves the problem of determining whether or not nondirective group play therapy could be applied to physically handicapped children.

They held 13 meetings with five physically handicapped children all of whom had at least one emotional problem. The children's teachers and parents made an essay-type report on the child's problems at the start of the program. At the last meeting similar reports were filled out again.

In a verbatim account the authors recount the play of a hemophiliac child. This boy would pretend to cut the therapist's fingers to cause him to bleed to death. The imaginary blood was collected in glass jars placed around the therapist.

The investigators found three of the children showed considerable observed improvement in both the home and the school. One child made slight reported gains, and one showed no improvement. They concluded that "the nondirective play-group offers an ideal setting for the self-solution for a particular type of emotional problem; namely, those stemming from the specific disability of the physically handicapped child" [13, p. 214].

The investigators themselves realize their work has been conducted at a very gross level. They point out several weaknesses of the project. Among these weaknesses are the lack of quantitative material. There were no pre- and posttherapy tests. There was no follow-up study to see if the indicated gains were temporary or cyclical in nature. Nor was a control group utilized to demonstrate more clearly that the play situation was the critical factor.

The recognition of such lacks is a healthy sign. It suggests an awareness of the possibility of improving future research work in play therapy by more rigorous procedure.

Nondirective play therapy in the handling of race conflicts. The effectiveness of play therapy for small groups of children who had difficulty adjusting to other children was the primary purpose of an investigation by Axline, "Play Therapy and Race Conflict in Young Children" [3]. She selected four groups of four children, two boys and two girls, who were either withdrawn or aggressively antisocial.

Each group met once a week for ten meetings. After the tenth meeting the children were mixed for five additional meetings. Axline [3] found that the Negro girls were accepted by the group after the seventh meeting. This new acceptance was carried over into the classroom. In all groups, "There was a tendency to participate in the group meetings with an awareness of the rights of others" [3, pp. 309–310]. The race problem was never an issue during the five mixed meetings.

This study is more provocative than it is definitive. The children selected were antisocial and not racially bigoted. It is quite possible that the figure of an habitual intergroup improvement in social relations was dressed in the false whiskers of lessening social antagonism. In that case Axline has given an old phenomenon a new name.

Nondirective play therapy in the treatment of reading disabilities. Since reading difficulty is frequently associated with emotional disturbances, it is not unexpected to find several research articles on the effectiveness of a therapeutic approach designed to relax and better adjust children who are retarded readers.

Axline [1] reported a study of 50 second-graders, listed as poor readers by their teachers, who were given a reading test. The 37 who received the lowest scores were placed in a special class. At the end of the semester, three and a half months later, intelligence and reading tests were administered.

There were 8 girls and 29 boys in the groups with Stanford-Binet IQ's ranging from 80–148. Unlike most remedial reading classes, these children had all their school work in one room with the same teacher. The reading problems were considered to be part of the whole child. "The children came first. The reading, writing, and arithmetic came secondly" [1, p. 65]. The children were given the opportunity for ample emotional expression. In accordance with the techniques of nondirective play therapy their feelings and attitudes were not only accepted but were also clarified. No remedial reading instruction per se was given.

Axline [1] found that 21 children gained more than the maturationally expected 3.5 in words. In the case of four subjects there was a noteworthy difference in the first and second IQ score. One subject's score was increased from 83 to 119.

This study would seem to indicate that nondirective therapeutic procedures are effective in building up a readiness to read in children.

In a later and briefer report Axline [6] studied three problem readers. Interestingly enough, only two of the children were poor readers, while the third child read too much. This child used books as a substitute for friends. All the children were above average in intelligence.

It was found that the feelings expressed in play brought out emotional problems that could easily account for the reading difficulties. Axline concluded, "Given the opportunity the child can and does help himself" [6, p. 161].

Neither of Axline's [1, 6] reports included experimental controls. Bills [9], working with 22 slow learners in the third grade utilized three 30-day periods of study. The first period was a control period in which all the children were tested with oral reading, silent reading, and Stanford-Binet tests. At the end of that period all the children were retested. They were also retested at the end of the second and third periods.

> The second period was the therapy period. The four children with the largest discrepancy between mental age and reading age were given nondirective play therapy. These children all had high IQ's. Four other children were selected whose IQ's were approximately average.
>
> The third period was used as a follow-up period in which the children were tested again.
>
> During the experiment, reading instruction was not remedial in nature and it was kept constant for all members of the class. Thus, a single group was compared with itself for three 30-day periods. Each child was his own control as the three periods were comparable in regard to reading experiences.
>
> The therapy group made a significantly greater gain in the therapy period than it did in the control period. The gains of the therapy group during both the second and third periods of the study were significantly greater than the gains during the first period of study.
>
> The gains in reading ability appeared immediately after therapy for some children and after a short period following therapy for others. The gains were found to be present six weeks (30 school days) after therapy had ended.
>
> Three judges agreed that five of the children had gained in emotional adjustment following play therapy. However, the design of the study did not permit conclusions as to the effect of maladjustment on children's reading ability.

To answer the question as to whether improved reading ability was due to improved personal adjustment Bills [10] conducted play therapy with well-adjusted readers. The design of the second study was similar to that of the first save that children were now selected for good adjustment by projective and objective personality tests.

He found reading gains were not significantly greater during the therapy period. So, it would appear that nondirective play therapy may improve reading in those children where emotional adjustment exists with the retardation. Consequently, play therapy is not necessarily the method of choice for all retarded readers as the reports of Axline [1, 6] might suggest.

Follow-up studies of nondirective play therapy. While it seems to have been demonstrated that play therapy is productive of personality improvement, it has not been shown whether the effects of therapy are permanent or temporary. Consequently, the value of follow-up studies cannot be denied.

Bills [9] found improved reading ability present six weeks after therapy. Axline (4) reports on an interesting study of a boy whose IQ was 65; upon retest it was 68. Six months after play therapy his IQ was 96. A year later it had gone up to 105. Part of this gain may be ascribed to test familiarization. However, the lasting effects of play therapy are again suggested.

In a long-range follow-up, Axline (7) selected 30 successful play-therapy case records. Of these, 22 subjects were available for follow-up study. Nineteen of the subjects were still successfully adjusted a year later, two were successfully adjusted three years later, and one five years after the original contacts. A follow-up of 24 of the 37 children used in previous research [1] was made five years later. Of this group originally designated as poor readers five were honor-roll students and four others had reading skills adequate for their grade placement.

The effects of play therapy then would seem to be lasting, particularly in the area of personality adjustment.

Summary and Critique

The principles and methods of nondirective play therapy are frequently presented as though they were firmly established. The assured manner of writing of many of the authors and the large-scale possibilities held before the reader, tend to make one believe that, at long last, "the way" has been found. Actually, this is not so. Indeed, it may not be the specific procedures of play therapy, per se, that effect the rather remarkable personality changes. The children may be benefiting from having someone constantly and consistently interested in their welfare. Those with experience in hospitals and institutions involving the mentally ill have noticed the unusually high percentage of cures attending any new treatment. They have reported that it is not the treatment method that effects the improvement, rather it is the increased interest taken in the patient. So, too, may it be with nondirective play therapy.

Axline [2] presented no experimental evidence to prove the worth of play therapy. A search of the literature reveals fewer than twenty published articles on the therapeutic uses of play therapy. To be admitted to the ranks of approved therapeutic methods nondirective play therapy needs more than enthusiasm, belief, and the shibboleth, "It works, if you only try it."

The greatest weakness of nondirective play therapy lies in this impetuous overlooking of the real need for a foundation in research. The most pressing need is for the employment of controls in play therapy. The personal adjustment of two equated groups should be assessed before therapy. The children in the group should then be randomly assigned to experimental and control groups. Upon termination of therapy all the children should be retested to determine the quantitative effectiveness of the play therapy technique in "helping children attain maturity."

Nondirective play therapy, while promising when evaluated subjectively, has been seen to have rather serious methodological lacks. One cannot concur with Kanner [17] that play work with children, while still in its beginning stages, "has come to stay," until play therapy has been established by objective means. In the long run, nondirective play therapy should stand or fall on the results of experimental studies investigating its effectiveness in relation to other procedures.

Received October 2, 1952.

References

1. Axline, Virginia. Nondirective therapy for poor readers. *J. consult. Psychol.,* 1947, **11,** 61–69.
2. Axline, Virginia. *Play therapy.* Boston: Houghton Mifflin, 1947.
3. Axline, Virginia. Play therapy and race conflict in young children. *J. abnorm. soc. Psychol.,* 1948, **43,** 300–310.
4. Axline, Virginia. Some observations of play therapy. *J. consult. Psychol.,* 1948, **12,** 209–216.
5. Axline, Virginia. Mental deficiency—symptom or disease? *J. consult. Psychol.,* 1949, **13,** 313–327.
6. Axline, Virginia. Play therapy—a way of understanding and helping reading problems. *Childh. Educ.,* 1949, **26,** 156–161.
7. Axline, Virginia. Play therapy experiences as described by child participants. *J. consult. Psychol.,* 1950, **14,** 53–63.
8. Baruch, Dorothy W., & Miller, H. Group and individual psychotherapy as an adjustment in the treatment of allergy. *J. consult. Psychol.,* 1946, **10,** 281–284.
9. Bills, R. E. Nondirective play therapy with retarded readers. *J. consult. Psychol.,* 1950, **14,** 140–149.
10. Bills, R. E. Play therapy with well-adjusted readers. *J. consult. Psychol.,* 1950, **14,** 246–249.
11. Bloomberg, C. M. An experiment in play therapy. *Childh. Educ.,* 1948, **25,** 177–180.
12. Cowen, E. L., & Cruickshank, W. M. Group therapy with physically handicapped children: II. Evaluation. *J. educ. Psychol.,* 1948, **39,** 281–297.
13. Cruickshank, W. M., & Cowen, E. L. Group therapy with physically handicapped children: I. Report of study. *J. educ. Psychol.,* 1948, **39,** 193–215.
14. Dorfman, Elaine. Play therapy. In C. R. Rogers, *Client-centered therapy.* Boston: Houghton Mifflin, 1951. Pp. 235–277.
15. Finke, Helene. Changes in the expression of emotionalized attitudes in six cases of play therapy. Unpublished master's thesis, Univer. of Chicago, 1947.
16. Fleming, Louise, & Snyder, W. U. Social and personal changes following nondirective group play therapy. *Amer. J. Orthopsychiat.,* 1947, **17,** 101–116.
17. Kanner, L. Play investigation and play treatment of children's behavior disorders. *J. Pediat.,* 1940, **17,** 533–546.
18. Landisberg, Selma, & Snyder, W. U. Nondirective play therapy. *J. clin. Psychol.,* 1946, **2,** 203–214.
19. Lebo, D. The relationship of response categories in play therapy to chronological age. *Child Psychiat.,* 1952, **2,** 330–336.
20. Miller, H., & Baruch, Dorothy W. Psychological dynamics in allergic patients as shown in group and individual psychotherapy. *J. consult. Psychol.,* 1948, **12,** 111–115.
21. Rogers, C. R. *Counseling and psychotherapy.* Boston: Houghton Mifflin, 1942.
22. Rogers, C. R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951.
23. Rousseau, J-J. *Emile.* New York: Dutton, 1925.

Journal of Consulting Psychology
Vol. 17, No. 3, 1953

The Relative Effectiveness of Larger Units Used in Interview Analysis[1]

John E. Muthard
Ohio State University[2]

This study investigated the relative effectiveness for studying client and counselor interaction of several larger units used in the analysis of transcribed interviews. This basic methodological question needs to be resolved so that future interview research can use the optimum unit, and so that findings from different studies can be compared and integrated. In addition, the study was concerned with the extent to which the units studied enable one to determine whether the different roles assumed by the client and counselor during an interview affect their respective behavior.

Protocol research has used a wide range of units to group client and counselor statements for the purpose of analysis. Early investigators in this area were principally concerned with studying single speech units, e.g., single remark and idea units. More recently, researchers have become more concerned with client and counselor behaviors in larger segments of the interview. The larger units used have been: the whole interview, discussion topics, and fractions of the interview or interview series [4, 5]. Larger units have several advantages over single speeches. They (*a*) have been found to be more reliably rated, (*b*) permit the study of meaningful structures in the interview which would otherwise be ignored, and (*c*) permit the study of delayed effects of both client and counselor responses on the progress of the interview. Comparing and integrating the findings arising from such a multiplicity of units is very difficult.

Procedures

Interview sample. To permit the widest possible generalization from this study, interviews were selected which represent not only different counselors, different clients, and different type problems, but also different institutions. Heretofore, each researcher has only used protocols obtained in his own institution. As may be surmised, different counseling centers attract different types of clients and problems and also often vary in the general counseling approach used. Thirty-six interviews were selected from a pool of 267 verbatim typescripts prepared from electrically recorded interviews at four universities: Chicago, Minnesota, Missouri, and Ohio State. The final sample included 12 client-counselor relationships. There were three sets of three consecutive interviews from each of the institutions. Wherever possible the cases selected included all the interviews in the series of contacts, but when this was not feasible—because of the length of the case—three consecutive interviews were selected from the case.

Units of the counseling interview studied. The larger units studied here include the discussion topic, fraction, and problem area. It was thought that the whole interview merited separate study since it has been found to have a definite pattern of organization.

The *discussion topic* represents all the conversation of both client and counselor about the same topic or subject. The division points for this unit were determined by the pooled judgments of three independent raters. After marking the division of the interviews independently, the judges met in conference to reconcile any differences. The character and merit of this unit is discussed at length by Robinson [3]. For a sample of nine interviews, two or all three judges agreed exactly or within the same remark for 83 per cent of the discussion-topic division points.

---

[1] Supported in part from funds granted to the Ohio State University by the Research Foundation for aid in fundamental research.

[2] Now at Emory University.

A *problem-area unit* consists of all the contiguous statements dealing with the same kind of problem. The initial step in determining these units is the same as for discussion topics. But since several successive topics may be aspects of the same type of problem, it is next necessary to group together those topics in the same class. The unit classification categories suggested by Robinson [4] were used, i.e., adjustment, skill, and special categories, to which was added a new class, test interpretation. Units in which the counselor told the client what the test meant and evoked little client participation were placed in this category. In the sample of interviews used to examine the reliability of demarcating this unit, two or all three judges agreed within the same remark 85 per cent of the time.

As a unit, the *fraction* comprises a definite percentage of the total remarks within a series of counseling interviews. For this research, quintiles of each series was the unit used. Some investigations using this unit are described by Raskin [2]. The reliability of determining its bounds is obviously quite high.

Dimensions of the counseling process used. To compare the three types of units studied, the behavior within each unit had to be considered. This was done by selecting dimensions which met the following criteria: reliability in rating, representativeness of the dimensions now used, and frequency of use in research. The concepts selected for use in this investigation were two counselor dimensions: amount of lead and counselor assumption of responsibility, and two client dimensions: client assumption of responsibility and working relationship [2]. Amount of lead refers to the degree which a counselor's remark seems to be ahead of the content of the client's last remark or brings pressure upon the client to adopt new ideas. Counselor assumption of responsibility is concerned with the amount of responsibility which the counselor assumes, which he permits the client to assume, or which he forces the client to assume. Working relationship refers to the degree that there is a mutual respect between the participants and the degree the client feels free to present his problems. In rating client assumption of responsibility, the judge considers the extent to which the client, in light of what the counselor is permitting him to do, takes an active role in clarifying his problem and participates in making plans for dealing with it.

Judging procedures. Persons participating in the ratings for this research were advanced graduate students in the student personnel program at the Ohio State University or had equivalent background or experience. They were trained in making whatever judgments they participated in. Three judges demarcated the units and rated the three global dimensions. Only two raters judged amount of lead, since this remark-by-remark rating can be done with high reliability. Precautions were taken to avoid introducing systematic variations in rating which might favor any of the three units studied.

Hypotheses. Four hypotheses were proposed as means of evaluating these units. Two of these dealt with reliability. They were: *Hypothesis 1:* There is a difference between the interrater reliability coefficients obtained for the ratings of the three types of units. *Hypothesis 2:* The units differ in the degree which the first half of each unit correlates with the second half. It was further predicted, in both instances, that ratings for problem-area units would be most reliable and those for quintile units would be least reliable.

The two other hypotheses considered a unit's capacity for distinguishing useful segments of the interview. That is, they examined the extent to which the units studied make variations within the interview discernible. *Hypothesis 3* was: There is a difference between the correlations obtained for ratings of two halves of a unit and those obtained for ratings of the last half of a unit and the contiguous half of the next unit. In the case of discussion-topic and problem-area units it was predicted that ratings of within-unit halves would correlate to a greater degree than those of between-unit halves. Fractions were predicted as not showing such a difference. *Hypothesis 4* was: Observable shifts in client and counselor behavior from each unit to its next adjacent unit differ when the difference scores[3] of the three units under study are compared. It was predicted that shifts would be greatest in the problem-area unit and least in fractional units.

---

[3] The numerical difference between mean ratings of adjacent units without regard to sign.

In each case the hypothesis was stated in the null form and tested.

Results

The reliability coefficients for the combined ratings indicated that the variables could be rated consistently enough to permit their use in research. For the three different-type units, the median reliability coefficient for the combined ratings on amount of lead was .93. Similarly, the median for counselor assumption of responsibility was .90, for client assumption of responsibility .72, and working relationship .71. Table 1 presents the complete reliability data for combined ratings.

Table 1
Reliability Coefficients for Combined Ratings
(Spearman-Brown prophecy formula used to take into account combined ratings)

Dimension	Discussion topic	Problem area	Fractions
Amount of lead	.93	.93	.96
Counselor assumption of responsibility	.90	.86	.94
Client assumption of responsibility	.72	.71	.73
Working relationship	.71	.69	.77

In terms of reliability, no one of the three units was found more advantageous than another. In only one of the twelve comparisons of correlations obtained from combined ratings was the obtained difference significant. In this instance, fractions were rated more reliably than problem-area units on the dimension counselor assumption of responsibility. Since such a difference might occur by chance during the course of 12 comparisons, the findings were interpreted as not providing sufficient evidence to reject the null hypothesis.

Neither could the hypothesis that the units do not differ in the degree which the first half of each unit correlates with the second half be rejected. The largest critical ratio obtained was .96 which failed to meet the fiducial limits set. Table 2 presents the correlation coefficients obtained in testing this hypothesis.

Table 2
Correlation Coefficients for Mean Ratings of Within-Unit Halves

Type Unit	Dimensions Amount of Lead	Working Relationship
Problem area	.79	.90
Discussion topic	.75	.89
Fractions	.72	.86

Although these findings concerning reliability do not support the alternatives proposed, they are amenable to explanation. Since the 36 interviews were divided into 60 fractional, 106 problem-area, and 174 discussion-topic units, it is apparent that fractions of the interview were generally three times as long as discussion topics and almost twice as long as problem-area units. Thus, the reliability for rating fractions was enhanced, since in rating or measurement, the larger the sample, the higher the reliability usually is for judging.

Another important criterion of the effectiveness of an interview unit for studying client and counselor behavior is the degree with which it makes discernible variation from one unit to another within an interview or series of interviews, that is, the extent to which it enables one to observe the variety of behaviors incorporated within an interview.

Hypothesis 3 can be rejected at the .001 level of confidence for both problem-area and discussion-topic units on both of the variables tested. In the comparisons involving these two units, the correlations for within-unit halves were always significantly greater than those for between-unit halves. However, there was not a significant difference for the fractional units studied. These findings are shown in Tables 3 and 4.

A test of the sensitivity to variations in client and counselor behavior of the three types of

Table 3
Significance of the Differences between Correlations of Amount of Lead Ratings for Within-Unit Halves and Between-Unit Halves

Type unit	*N*	Correlations for within-unit halves	Correlations for between-unit halves	Critical ratio
Discussion topic	147	.75	.49	3.71*
Problem area	81	.79	.35	4.40*
Fractions	48	.72	.85	1.65

*Significant at the .001 level using one-tail test.

Table 4

Significance of the Difference between Correlations of Working Relationship Ratings for Within-Unit Halves and Between-Unit Halves

Type unit	N	Correlations for within-unit halves	Correlations for between-unit halves	Critical ratio
Discussion topic	147	.89	.66	5.34*
Problem area	81	.90	.57	5.15*
Fraction	48	.86	.86	0.00

*Significant at the .001 level of confidence using the one-tail test.

Table 5

Significance of the Difference between Sets of Difference Scores

(Based on analysis by Mann-Whitney U test [1])

	Unit Comparisons		
Dimension	Problem area vs. fraction	Topic vs. fraction	Problem area vs. topic
	x*	x	x
Amount of lead	5.31†	4.81†	1.50
Counselor assumption of responsibility	3.41†	3.30†	.76
Client assumption of responsibility	2.62†	1.48	1.64‡
Working relationship	2.67†	1.81‡	1.21

* Similar to the critical ratio.

† With an $x > 1.96$ the null hypothesis can be rejected at the .01 level of confidence.

‡ With an $x > 1.64$ the null hypothesis can be rejected at the .05 level of confidence.

units was made with hypothesis 4. The results obtained from examination of this hypothesis are shown in Table 5.

On all four of the variables used in this investigation, problem-area unit ratings yielded difference scores which were significantly larger, at better than the .01 level, than those obtained for quintiles of the interview series. As Table 5 indicates, there is a significant difference for three of the dimensions studied when discussion topics and fractional units are compared. However, when problem-area and discussion-topic units were compared, only one difference was barely significant. In general then, both problem area and discussion topic produce greater variation from unit to unit than does the use of fractions. We have little basis for determining whether the problem area or discussion topic is better than the other. It should be noted, however, that in this comparison all the differences were in the direction of greater sensitivity for problem-area units and that 58 per cent of the difference scores for these two sets were identical.

An analysis of client and counselor behavior during the course of a "typical" interview suggested that one of the factors associated with behavior changes between successive problem-area or discussion-topic units is changing-role[4] behavior. That is, the variations within the conference may be a function of the demands and expectations which the client and counselor perceive in the counseling relationship for different types of problems and topics. This research suggests that a more systematic evaluation of client and counselor roles may lead to a better understanding of the dynamics of the interview, i.e., the interaction phenomena and variations in client and counselor behavior.

Discussion

We have found that discussion-topic and problem-area units when compared to a fractional unit (*a*) show greater sensitivity to variations in behavior during the course of a conference and (*b*) bring together client and counselor responses which reflect related behaviors and attitudes to a greater degree. Thus, it seems imperative that any unit used in interview typescript research be based on a psychologically meaningful, organizing principle. That is, units should bring together behaviors and attitudes which are of the same pattern, and partition those segments of the interview which are dissimilar psychologically. Discussion-topic and problem-area units met this standard.

The thorough analysis of a single "typical" interview and general impressions from reading over 200 interview protocols suggested the possibility of counselor or client roles as a basis for unit divisions. A study of roles assumed in the counseling interview may lead to further understanding of the dynamics of the interview.

Since this is one of the first studies which use counseling interview records from several institutions, it was felt that there would be some value in investigating whether the schools differ on the dimensions studied. Since the data secured represented a limited sampling — three different clients in nine interviews— from each campus, the findings must be considered very tentative. Nonetheless, some significant differences were found in problem areas discussed and in the pattern of techniques used. The findings therefore suggest that, in making studies which are concerned with the interaction of client and counselor, the differential distribution of problem types and the counseling approaches at each of the schools studied should be considered.

---

[4] For our purposes, Sargent's [6] definition of role seemed useful. He says, "A person's role is a pattern or type of social behavior which seems situationally appropriate to him in terms of the demands and expectations of those in his group."

Summary and Conclusions

This study examined the relative effectiveness of three larger units used in interview typescript analysis. The three criteria used to examine effectiveness were: (*a*) reliability, (*b*) sensitivity, and (*c*) degree to which the unit brings together related material and partitions that which is less related.

It was found that discussion-topic, problem-area, and fractional units could all be rated with acceptable reliability and that the rating reliabilities obtained were not significantly different. On the other two criteria, it was found that both problem-area and discussion-topic units were more effective than quintiles of the interview series. Problem-area and discussion-topic units were not found to differ on these criteria.

Examination of the protocols suggested that changing client and counselor roles may be important factors associated with variability within the conference and are directly associated with topic and problem areas. Thus, the discussion-topic and problem-area units provide an opportunity to study such roles, but quintiles of the interview series do not.

Received October 20, 1952.

References

1. Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.,* 1947, **18,** 50–60.
2. Raskin, N. J. An analysis of six parallel studies of the therapeutic process. *J. consult Psychol.,* 1949, **13,** 206–220.
3. Robinson, F. P. The unit in interview analysis. *Educ. psychol. Measmt,* 1949, **9,** 709–716.
4. Robinson, F. P. *Principles and procedures in student counseling.* New York: Harper, 1950.
5. Rogers, C. R. (Ed.) *Client-centered therapy* Boston: Houghton Mifflin, 1951.
6. Sargent, S. S. Role and ego in psychology. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads.* New York: Harper, 1951. Pp. 355–370.

Journal of Consulting Psychology
Vol. 17, No. 3, 1953

On the Supposed Behavioral Correlates of an "Eye" Content Response on the Rorschach

Michael Wertheimer
Wesleyan University[1]

Because it is clinically of the utmost importance to know as much as possible about the individual case, any kind of information, whatever its source, is often considered to be of great value. Thus one finds that in practice many correlates for test behavior are used whose empirical foundations are not optimal. If a practitioner has found in his experience that *A* tends to go with *B*, he may thenceforth attribute the characteristic *B* to patients who give him *A* on a test, and soon it may become standard diagnostic practice to consider *A* a "sign" of *B*.

This situation is of course inevitable in a field so complex as that of personality diagnosis, where there is such a great need for information. Much research is going on at present to answer this demand. Considerable progress has already been made in the revision of theory, such as in the areas of color shock, experience balance, and diagnosis of organicity, to mention only a few. Empirical devices have recently also been brought to bear on the general area of the content of responses on the Rorschach test.

The present study attempts validation of a representative content item. The item was chosen with regard to two criteria: first, it should be a relatively well-accepted interpretation; second, it should be one which is relatively easy to test objectively.

Lindner [2] writes of Card IV, D1, "The shading characteristics of this region encourage the perception of a 'pair of eyes.' Nevertheless persons with strong paranoid tendencies, and those whose symptom complexes include ideas of influence or reference invariably note the 'eyes' and express the opinion that they are 'looking at me.'" Elsewhere he states, "Persons with strongly paranoidal tendencies, however, and whose symptom complexes include ideas of influence or reference, deliberately note the 'eyes,' abstracting them from the head of the insect or animal and adding the impression that 'they are looking at me.' Such responses appear with similar frequency from strongly inferiority-convicted persons. Among these such a response probably represents a projection of the idea of being under the close scrutiny of the examination" [3, p. 83]. Hertz writes, "Ideas of influence or reference may be detected by the persistence of eyes" [1, p. 15], and Schafer [4], in reference to paranoids, says that some of these patients emphasize "eyes." While analyzing a Rorschach record, he writes, "On Card IV, under the impact of anxiety, his integrative ability is so impaired that he is restricted to the eyes of what is frequently seen as a full animal head (lower middle *D*), and misses the popular *W* (animal skin) altogether. On Card IX, another frequently disturbing card, precisely the same thing happens: after a long delay he makes out a pair of eyes, one of which is frequently integrated into a full animal head. The choice of content 'eyes' can be considered a further indication of his paranoid trend . . ." [4, p. 155]. Courses in practical Rorschach technique often consider *eyes* as an important, or even an unmistakable sign of paranoid trends; and many workers in the field use it as an important indicator in the practical everyday work of diagnosis.

These statements about the relationship between paranoid trends and *eyes* content seem to satisfy the two criteria of being well accepted and relatively easy to test.

---

[1] This research was performed while the writer was USPHS Clinical Psychologist at the Worcester State Hospital.

Procedure

The files of the Psychological Laboratories at the Worcester State Hospital were used for the purpose of a statistical check. Altogether, 230 records were examined. An attempt was made to find 15 records of males and 15 of females in each diagnostic group. In some categories less than 30 records were available, so that the *N* was not always the same from group to group. The diagnostic categories used were: (*a*) paranoid psychoses including schizophrenia and other paranoid conditions; (*b*) schizophrenia, other types, which can also be called the schizo-affective group; (*c*) nonparanoid and nonaffective schizophrenics, including simple, hebephrenic, and catatonic schizophrenics; (*d*) manic-depressive patients, whether manic, hypomanic, depressed, or mixed; (*e*) organic patients, including alcoholics, epileptics, and those with disturbances of the central nervous system; (*f*) neurotics, independent of further classification; (*g*) a group without psychosis, diagnosed simple adult maladjustment; and finally (*h*) psychopaths.

For each patient, age, sex, psychiatric diagnosis, Stanford-Binet Vocabulary IQ and the total number of responses to the Rorschach were recorded. The record was then combed to find every occurrence of the word *eyes*, and the number of *eye* responses was recorded, with reference to whether they occurred in the free association or in the inquiry, and on which card they appeared. The original intent was also to record the detail which had given rise to the response, but the data on the protocols were usually insufficient to give clear indications of location. Records used were obtained in the period 1936 to 1952, from patients referred to the psychological laboratories by the staff of the hospital. All were obtained by trained psychologists, ranging in experience from intern to director of psychological research.

Table 1

The Incidence of Eye Responses on the Rorschach Cards

Card	Number of eye responses	Proportion
I	34	.09
II	29	.08
III	22	.06
IV	83	.22
V	17	.04
VI	37	.10
VII	26	.07
VIII	23	.06
IX	54	.14
X	52	.14
Total	377	1.00

Results

Table 1 shows the number of times that the word *eyes* occurred in response to the various cards. The table is based on the records obtained from the following groups: paranoid, schizophrenia other types, neurotics, and organics; 119 records in all. Only 14 female organic records were found in the files. Lindner's statement about Card IV that the "shading characteristics . . . encourage the perception of a 'pair of eyes' " is consistent with the finding that Card IV produces the most frequent mention of eyes. It must be mentioned here that the number entered in the table is the number of times the word *eyes* appears in the free association, plus the number of times it occurs in the inquiry, independent of context.

In order to get a more adequate picture of the "importance" of eyes in a record, each record was corrected for the total number of responses. An *eye index*, the number of times the word *eyes* appeared in the record, divided by the total *R* of the record, was obtained. This eye index was plotted against IQ, in order to see whether there was any relationship between the two. Lumping the 230 records into four groups, Table 2 was obtained. Means and medians for each group are given, and both show the same trend. The great difference between the means and the medians is due to the extreme skewness of the distribution of the eye index; 91 of the 230 records

Table 2

Eye Index as a Function of IQ

IQ Range	*N*	Mean eye index	Median eye index
60– 79	27	.197	.085
80– 99	68	.115	.030
100–119	82	.158	.090
120–139	53	.166	.100

included not a single *eye* response, and therefore almost 40% of the records had an eye index of 0.00. There is some indication in both statistics that the middle range of intelligence seems on the average to be less productive of *eye* responses on the Rorschach than do the extremes.

To see whether the production of *eyes* on the Rorschach is a function of age, the eye index was plotted against chronological age. Such a plot, when age is broken down into five groups, gives Table 3. Since the means give an adequate picture of the relationship, medians are not given in the table, although the distributions are once more markedly skewed. The increase in the number of *eye* responses with increasing age is clear. What it signifies is open to speculation.

Table 3

Eye Index as a Function of Age

Age range	*N*	Mean eye index
10–19	38	.100
20–29	73	.130
30–39	64	.135
40–49	39	.187
50–59	16	.218

To turn to the relation between *eye* response and diagnosis, Table 4 shows the proportion of subjects who reported *eyes* on Card IV, divided into the various diagnostic groups.

Table 4

The Relation between Incidence of Eye Responses and Diagnosis

Diagnostic group	Proportion of *S*'s reporting eyes on card IV: In free assoc.	Proportion of *S*'s reporting eyes on card IV: In inquiry	*N*
Simple adult maladjustment	.03	.28	29
Schizophrenia, other types	.10	.37	30
Simple, hebephrenic, and catatonic schizophrenics	.13	.33	30
Psychopaths	.11	.38	27
Organics	.21	.29	29
Neurotics	.13	.43	30
Paranoids	.17	.43	30
Manic-depressives	.32	.44	25

This result is somewhat discouraging in that we would expect the paranoid patients to stand out considerably above the others. Lindner's strong statement that paranoids "invariably" note the eyes is not consistent with this finding. Paranoid patients do, however, report eyes relatively frequently compared with other diagnostic groups.

Table 5 shows the mean eye index for the various diagnostic groups; the result here is more encouraging, but it must be remembered that this is eye index for all cards, not specifically for Card IV alone.

Table 5

Eye Index As a Function of Diagnostic Group

Diagnostic group	Mean eye index	*N*
Paranoids	.178	30
Neurotics	.164	30
Manic-depressives	.142	25
Schizophrenia, other types	.116	30
Simple, hebephrenic, and catatonic schizophrenics	.111	30
Psychopaths	.108	27
Simple adult maladjustment	.105	29
Organics	.097	29

It is clear from this analysis that a response of *eyes* on Card IV, or a high eye index in the record, cannot be used automatically to infer paranoid ideation in a subject who produces such a record.

It may well be objected at this point that psychiatric diagnosis is far from infallible, and that there may be features in a case which are more important than paranoid trends, and which force the psychiatrist to make a diagnosis of something other than a paranoid disease even though a patient may show clear suspiciousness. In an attempt to answer the critique, a further analysis was performed. A file of Keysort Cards, one for each patient examined, was available in the office of the Director of Psychological Research at the Worcester State Hospital. These cards include, among other information, a code for presence or absence of suspiciousness in the case history of the patient. Table 6 shows the analysis of the suspicious and nonsuspicious groups into records which showed at least one response of *eyes* on Card IV as against those which did not; also those records in which at least one

Table 6

The Number and Proportion of Patients Showing Eye Responses either on Card IV or on the Entire Record, as a Function of Whether Their Symptoms Included "Suspiciousness" or not.

Sign	Not susp.	Suspicious	Total	Sign	Not susp.	Suspicious	Total
Eyes on Card IV	51 (34%)	35 (45%)	86	Eyes in record	86 (57%)	53 (68%)	139
No eyes on Card IV	101 (66%)	43 (55%)	144	No eyes in record	66 (43%)	25 (32%)	91
All cases	152	78	230	All cases	152	78	230

Table 7

The Number and Proportion of Patients Showing Two or More, or Three or More Eye Responses in the Entire Record, as a Function of Whether Their Symptoms Included "Suspiciousness" or not.

Sign	Not susp.	Suspicious	Total	Sign	Not susp.	Suspicious	Total
2 or more eyes in record	70 (46%)	38 (49%)	108	3 or more eyes in record	55 (37%)	33 (41%)	88
Less than 2 in record	82 (54%)	40 (51%)	122	Less than 3 in record	97 (63%)	45 (59%)	142
All cases	152	78	230	All cases	152	78	230

response of *eyes* occurred on any card as against those on which *eyes* were never mentioned.

Chi square was calculated for both of these two-by-two tables; in each case the probability that such distributions may have been drawn from a population in which there was no difference in the production of *eyes* between the suspicious and the nonsuspicious groups was just under .10. Chi squares for the two tables were 2.82 and 2.79 respectively; chi square corresponding to a p of .10 is 2.72; for $p = .05$, it is 3.84. By the generally accepted criteria of statistical significance, therefore, it is suggested that neither production of *eyes* on Card IV nor production of *eyes* anywhere in the record differentiates between the suspicious and the nonsuspicious group.

The mean eye index of the two groups, .156 for the suspicious and .106 for the nonsuspicious groups, is suggestive of a difference between them.

One other criterion of paranoid trends, that of Hertz, remains to be tested. It is claimed that the persistence of *eyes* in a record is the determining criterion of suspiciousness in the patient. For analytical purposes, Table 7 gives the number of records showing two or more, and three or more *eye* responses, for both suspicious and nonsuspicious patients. This might be one approach to the definition of "persisence." Chi squares calculated on both of these come out with probabilities of .70 and .40 of obtaining such results by chance, suggesting that at least for the present sample, and with persistence defined as recurrence, there is no difference in the persistence of *eyes* between suspicious and nonsuspicious patients.

A few other relationships were suggested by the data. Females tend to have more *eye* responses in their records than do males, independent of diagnostic group; but this result was not statistically significant. Also, the suspicious group tended to have fewer responses in the whole record than the nonsuspicious, but again this difference is not statistically significant. The median values of R for the two groups were 15.5 and 16.6 respectively, the means 18.7 and 19.2.

At best, then, we can say that there might be some statistical trends in support of the statements given in the introduction, but none of these trends reaches the criteria of statistical significance which are currently in use. Even if clearly significant statistical differences were found, this would still be a far cry from the "invariable" type of relationship, the existence of which is sometimes asserted in works on Rorschach content analysis. On the other hand, it must be said in all fairness that such statements of one-to-one relationships are the ones which lend themselves most readily to quick empirical check.

Discussion

Before further generalizations are made on the basis of these data, certain rather serious considerations must be mentioned. In the tabulations, a completely objective attitude was used; every single time that the word *eye* or *eyes* occurred, a check-mark was made. This independence from context, which the hypothesis required, has both great advantages and grave disadvantages. It makes it impossible for bias on the part of the investigator in any way to affect the results; it is simple; and it permits easy statistical analysis. On the other hand, such an atomistic approach to the hypothesis may strike the practicing clinician as rather naive, in that it may be considered obviously incorrect to divorce the word *eye* completely from its context. However, this is a direct translation, into experimental terms, of the "cookbook" approach to Rorschach interpretation. Although the authors of such books may often pay lip service to the importance of the context of the response, still, in the usual form of presentation, a one-to-one correlation of response to diagnostic significance is either implied or made necessary as a working hypothesis for the student. It may be argued that, for didactic purposes, such oversimplification is necessary and justified. It may be, however, (and this is the point of this study) that this method of teaching destroys the essence, and therefore the validity, of the technique. The findings of this study certainly seem to point to such a possibility. Therefore this study aims at the evaluation of a method of teaching the Rorschach, and may not have any immediate relevance to the interpretation of the Rorschach by an experienced clinician, although some implications may be drawn for Rorschach interpretation in general.

It may well be that *eyes* in a record of a certain configuration may be highly related with paranoid trends, whereas another gestalt results in no predictive significance of *eyes* at all, i.e., there is a world of difference between the following response in the inquiry, "Well, this is his head; here are the ears, here is the nose, and here are the eyes and mouth," and on the other hand, the response in the free association: "All I see on this are two eyes, looking straight out from the gloom."[2] A test in which, for example, a number of experienced Rorschach workers would be asked to rate each *eye* response should eventually be undertaken, but would be a test of a different hypothesis than the one tested here.

The present tendency towards objectification and towards clear-cut statements of relationships places more and more Rorschach materials within reach of the experimenter's techniques. On the other hand, the same type of approach, in making extremely explicit statements on the basis of clinical experience, may sometimes unavoidably carry the connotation of extensive empirical justification. Such an implication regarding a statement for which there is as yet insufficient data may blind workers in the field to the need for objective research.

If one considers the present study as an adequate test of the hypothesis that the *eye* response, as such, tends to go with paranoid trends, certain further ideas offer themselves for consideration.

First of all, this study suggests that the general acceptance of an interpretation is by no means an index of its empirical validity. Perhaps it is precisely the most commonly accepted interpretations which are the ones most critically in need of objective analysis.

Secondly, while the testing of the present hypothesis presents some difficulties, other relationships between test results and behavior which have become traditional in the field of the Rorschach may be even more difficult to test. Such relationships may be less definite with respect both to test criteria and to behavior correlates. It is less difficult to get objective agreement on the presence or absence of a word in a record than on items such as $F + \%$; on the behavior end, it is easier to determine the presence of paranoid trends than it is to determine the presence or absence of strong hostile desires which are well controlled and not acted out.

---

[2] There were less than ten records in the entire 230 which included responses as extreme as those of the second type mentioned, and by no means all of these occurred in records of individuals who had a paranoid disease or whose case histories gave evidence of suspiciousness.

Until further research is completed and the results made available to the practicing Rorschach worker, perhaps it would not be amiss to emphasize caution in the interpretation of Rorschach records. Whereas the general tendency in the field today is to make as ominous an interpretation as possible, at least in a mental hospital, perhaps the other extreme is more called for, if indeed one extreme is to be preferred to the other. Since most of the indices in use at present are probably not nearly as accurate differentiators as we would have them be, we would be more likely to help a patient by making definite interpretations only of those few things on which there is good evidence, and for the remainder, by being aware of the many possible alternative implications (some perhaps pointing to assets and to health) and the good-sized margin of error.

This study, then, indicates that particular caution should be exercised in making one-to-one interpretations, and leads to the suggestion that attempts to find objective Rorschach correlates for different behavior patterns would profit more by investigating the configurations and interactions of various indices, than by attempting to assign individual and invariable meaning to the indices.

Summary

Two hundred thirty Rorschach records from different diagnostic groups were examined for the frequency and incidence of *eye* content responses. Although much of the literature states that an *eye* content response to the Rorschach is indicative of paranoid or suspicious trends in the subject, this statement is not borne out by the present data. Paranoid patients did not produce significantly more *eye* content responses than did other diagnostic groups, nor did patients whose symptoms include suspiciousness produce more *eye* content responses than did patients whose symptoms did not include suspiciousness. Caution is urged in the use of one-to-one behavior correlates of Rorschach signs.

Received October 20, 1952.

References

1. Hertz, Marguerite R. Suicidal configurations in Rorschach records. *Rorschach Res. Exch.*, 1948, **12**, 3–58.
2. Lindner, R. M. Analysis of the Rorschach test by content. *J. clin. Psychopathol.*, 1947, **8**, 707–719.
3. Lindner, R. M. The content analysis of the Rorschach protocol. In L. E. Abt & L. Bellak (Eds.), *Projective psychology*. New York: Knopf, 1950.
4. Schafer, R. *Clinical application of psychological tests.* New York: International Univer. Press, 1948.

Journal of Consulting Psychology
Vol. 17, No. 3, 1953

The Levy Movement Test: Suggestions for Scoring and Relationship to Rorschach Movement Responses

Robert M. Allen, Charles D. Ray, and Robert C. Poole[1]
University of Miami

At the present time there is little agreement in scoring movement determinants imposed by the subject on the Rorschach ink blots [1, p. 18]. Another difficulty related to the movement response results from the complexity of the blots, i.e., the plates do not present a homogeneous stimulus field but are composed of varying degrees of form, shading, chromatic, and achromatic variables affecting perception. A third problem springs from the relatively nondirected nature of the orientation of the subject to the Rorschach task.

Problem

In studying movement apart from other variables found in the Rorschach test situation, it would appear paramount to limit the task to less complex stimuli. A new and simpler instrument, the Levy Movement Cards (LMC),[2] draws its value from being less subject to these difficulties. This new set of stimuli has been developed for the express purpose of studying correlates of the movement responses in the Rorschach. The plates are achromatic, bilaterally symmetrical, finger paintings showing humanlike figures in each half of the card. Color is absent, ostensibly to simplify the stimulus and to hold to a possible minimum any determinant-complexes arising from the presence of hue which might inhibit or facilitate movement-response tendencies. As a further measure in orienting the task to the perception of movement, the subject is given the following directions when presented with the LMC: "I am going to show you some movement blots. These show people doing something. Look at the cards in the position in which I give them to you,[3] and tell me what the people are doing. When you are through, place the card face down on the table and I will give you the next one. Do you have any questions?"

In this paper the authors will (*a*) define and illustrate a simplified scoring scheme, and (*b*) compare the different types of movement concepts given to the LMC with those elicited by the Rorschach. In order to compare the movement data obtained from these tests a scoring system was devised for the Levy movement responses to parallel the Rorschach. The system proposed by Levy [6] and Zubin [9] was too complex for handling such data. In place of their numerically detailed scales for rating compliance, movement, content, and psychological nature of the response, the revised scoring schedule includes the following elements:

1. Empathy: the subject's acceptance of the task. *X—Rej*—rejection of the card, no response given. *X—*response(s) other than movement given. *N-Pr*—movement response not spontaneous. *Pr*—movement response present and spontaneous.
2. Movement: concept seen in motion. *F*—form or other nonmovement response given. *OM*—inanimate movement response (Klopfer's *m*) *AM*—whole animal seen in motion (Klopfer's *FM*) *AdM*—detail of animal in movement. *HM*—whole human seen in motion (Klopfer's *M*). *HdM*—detail of human in movement.
3. Forces: contributions made by external (*E*) and internal (*I*) forces in the movement response as seen by the subject; also refers to the level of

---

[1] The authors are indebted to Dr. Carl Williams, Department of Psychology, University of Miami, for his valuable assistance with the statistical phase of the paper.

[2] The Levy Movement Cards were supplied by the Psychological Corporation for research purposes. This kindness is greatly appreciated.

[3] One card, VI, is presented in three different positions consecutively.

Table 1
Summary of Productivity and Movement Responses for LMC and Rorschach Tests

Measure	Total Responses		Human Movement		Animal Movement		Other Movement		Total Movement	
	LMC	R	LMC	R	LMC	R	LMC	R	LMC	R
Raw Mean	13.1	26.2	8.8	3.7	.6	3.8	2	.6	11.4	7.7
Raw Median	11	21	10	2	0	3	1	0	10	7
Per cent Mean			78	15	3	14	11	2	92	31
Per cent Median			90	10	0	12	8	0	92	31.6

movement energy imposed in the concept. These may be equated with Rorschach flexor and extensor movement.

E—mechanical, inanimate movement: falling, floating, balancing (by gravity), being pushed, blown apart, etc.

E/I—static, positional animal or human movement: standing, leaning, poised, holding something, talking, sitting, breathing, etc.

EI—active animal or human movement: walking, bending, bird flying, fish swimming, etc.

I/E — energetic animal or human movement: dancing, diving, swimming, running, jumping, fighting, walking or standing on hands, intercourse, etc.

I—uncontrolled, fantastic, inanimate movement: exploding, etc. Movement must be forced from the internal, otherwise score *E*.

4. Content: standard scoring as in the Rorschach test.

5. Popularity and originality: apply Rorschach standards.

Example: Levy Card VI, "Looks like two people in bathing suits, diving into a pool of water." Movement response is spontaneous, *Pr;* human movement seen, *HM;* movement is energetic, *I/E;* clothing and water mentioned, *Clo, N*. Scored: *Pr HM I/E Clo, N;* Rorschach equivalent: *W M H*.

Method

In a preliminary study with 25 college students the above proposed scoring scheme proved to be adequate for representing and comparing the nature, content, and level of empathy of the movement responses in the LMC and Rorschach protocols. Each subject was administered the two tests by the individual method and by the same tester. In every instance the Rorschach preceded the LMC Test. This procedure was established because the LMC instructions definitely set the subject to perceive human movement. Had these tests been given in the AB-BA order, the LMC instructions could conceivably influence the spontaneous production of movement responses in the Rorschach plates. The results could be unnecessarily confounded by this variable.

Results and Discussion

An examination of the distributions of the responses reveals the nonparametric nature of the data and the consequent inability to assume normality. Table 1 presents movement-response means and medians based on the absolute raw number of responses and the ratio of movement to total responses (percentage) given by each subject on each test separately. The total productivity for each test is also indicated. The number of raw mean and median total responses for the Rorschach is double that for the LMC. Only three subjects have more than 15 responses in the latter test, none less than nine; the limited range of 9 to 14 responses includes 84 per cent of the subjects. This is not surprising in view of the rigidity of the LMC directions as compared with the relative freedom of the Rorschach instructions. The LMC directions seem to imply that giving a human movement percept relieves the subject of the need for further response to the card. This is not so with the Rorschach test.

Again Table 1 reveals a dramatic story: the movement story. For this selected population, as a group, movement was a major determinant in 92 per cent of the LMC responses and in 31 per cent of the Rorschach associations. As indicated above, this should be the expected direction of variance because of the specific movement-perception set imparted by the LMC instructions. This is supported by an analysis of the movement responses into three types: human, animal, and other. Using the median percentage data (to mitigate the distorting effects of extreme ratios), it may be seen from Table 1 that LMC human-movement percentage greatly exceeds Rorschach movement, 90 per cent to 10 per cent human-movement associations, respectively. In keeping with the attitudinal set, animal responses are minimal in the LMC while the Rorschach elicits a median of 12 per cent *FM*. Again one must turn to the LMC drawings for a possible explanation—animal engrams are not as readily elicited because of the instructions and the easily recognized humanlike quality of the finger paintings, much more so than the nebulous ink blots. For future research, then, there remains the problem of evaluating the rationale behind the animal responses in the LMC Test which represents a definite departure from specific instructions.

Other movement concepts are consistently higher for LMC than for the Rorschach. An analysis of the contents of the subject's responses again points to the structure of the individual LMC plates which determines such associations as "wind blowing," "rain falling," and "whirling" movements — concepts which are readily engendered by the chiarascuro tones of the finger paintings which are suggestive of natural phenomena. To recapitulate: Table 1 discloses (*a*) data related to total productivity for the LMC and Rorschach tests, and (*b*) an analysis of the three types of movement associations into human, animal, and other.

Table 2

Correlation between LMC and Rorschach Movement and Total Responses

Response	Rho
Human movement	+.20
Animal movement	—.07
Other movement	—.08
Total movement	+.24
Total responses	+.50

The question which flows logically from the findings presented in Table 1 is: what is the relationship between movement associations of the LMC and Rorschach tests? Two nonparametric statistics are utilized in order to approach the answer to this question—rank correlation, rho; and Wilcoxon's method for individual comparisons by the ranking method [2, 4, 7, 8]. Table 2 discloses the rho coefficients between the types of movement responses obtained on the LMC and Rorschach tests and the correlation between the total number of responses for both tests. None of these values is significant for estimating a definite relationship between the two variables [3, p. 314]. One cannot state, if such an hypothesis entered into the considerations of the LMC proponents, that a high number of movement associations in the LMC Test is necessarily accompanied by increased movement responsiveness in the Rorschach ink blots, or vice versa. The same hypothesis is untenable for total productivity on both tests. These findings suggest further that there is no acceptable means of predicting, for the individuals within the study group, to what extent productivity and movement engrams in one test will be reflected in the other.

Table 3

Significance of LMC and Rorschach Movement Percentages

Response	Diff.	p
Human movement	63	$<.01$
Animal movement	11	$<.01$
Other movement	9	$<.02$
Total movement	61	$<.01$

An examination of the raw percentage distributions of the responses leaves the impression that, as a group, there are some significant differences in LMC and Rorschach responsiveness. The percentage data were analyzed for significance in accordance with Wilcoxon's method [7, 8].[4] The probabilities are presented in Table 3. The implications are clear: there are significant differences in movement-determined concepts for the group on the LMC and Rorschach tests. Human-, other-, and total-movement responses are elicited more prominently by the LMC; animal-movement associations are predominantly engendered by the Rorschach ink blots. These results support the data reported in the earlier portion of the paper.

[4] This is a nonparametric test based on rank order and is known as Wilcoxon's T test. No assumption of normality of distribution is made.

Conclusions

Can the LMC Test be employed as a means for checking the Rorschach movement responsiveness of a subject? It would appear that the present findings, with this group of subjects, militate against such an hypothesis. All that may be concluded is: (*a*) The Levy Movement Card Test elicits a significantly higher percentage of movement responses than does the Rorschach Psychodiagnostic Test. (*b*) The LMC Test is less productive of total responses than the Rorschach. (*c*) While the above two conclusions hold for this study population as a whole, prediction for an individual subject's total and specific type of movement responsiveness cannot be made. Thus, it is difficult to evaluate high or low productivity by a subject in one test in terms of productivity in the other test. (*d*) As a final opinion, pending futher research, the authors believe that the greater degree of structure in the LMC instructions and the nature of the drawings themselves could account for the results obtained. The authors suggest that much more work has to be done with the LMC Test before it may justifiably be employed as a check on Rorschach movement responsiveness.

Received September 22, 1952.

References

1. Allen, R. M. *Student's Rorschach manual.* Miami, Fla.: Dept. of Psychology, Univer. of Miami, 1950, mimeographed.
2. Friedman, M. The use of ranks to avoid the assumption of normality. *J. Amer. statist. Ass.*, 1937, **32**, 675–701.
3. Guilford, J. P. *Fundamental statistics in psychology and education.* (2nd Ed.) New York: McGraw-Hill, 1950.
4. Moses, L. E. Non-parametric statistics for psychological research. *Psychol. Bull.*, 1952, **49**, 122–143.
5. Rust, R. M. Some correlates of the movement response. *J. Pers.*, 1948, **16**, 369–401.
6. Rust, R. M. III. The Levy Movement Cards: EPA Round Table. *J. Pers.*, 1948, **17**, 153–156.
7. Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics Bull.*, 1945, **1**, 80–82.
8. Wilcoxon, F. Probability tables for individual comparisons by ranking methods. *Biometrics*, 1948, **3**, 119–122.
9. Zubin, J., & Young, K. M. *Manual of projective and cognate techniques.* Madison, Wis.: College Typing Co., Univer. of Wisconsin, 1948, multigraphed.

Journal of Consulting Psychology
Vol. 17, No. 3, 1953

Prognosis in Paranoid Schizophrenia on the Basis of the Rorschach[1]

David Grauer
Veterans Administration Hospital, Hines, Illinois

Previous published prognostic investigations of the Rorschach on schizophrenics have made no systematic attempt to isolate the paranoid subtype for special study. Evidence in the literature, however, suggests that the paranoid subtype is distinct from others and therefore merits consideration as a special group [2, 11, 14, 22, 24]. For this reason, the present study was confined to paranoid schizophrenics.

Several investigations of the Rorschach as a prognostic instrument in schizophrenia have been reported [8, 9, 13, 15, 16, 17, 19, 20, 21]. Each of the studies has discovered combinations of Rorchach signs which significantly distinguished improved from unimproved schizophrenics. There are, however, many inconsistencies and contradictions in the specific signs found to be prognostic, even among the investigations of the psychologist who has contributed most to this literature [19, 20, 21]. A recent review by Windle [25] summarized the results of many of these prognostic studies.

A definite explanation for the diversity of these findings cannot, of course, be given in the present state of our knowledge. Nevertheless, it is clear that a majority of the investigations are deficient in many respects. In the first place, prognostic signs were established on a purely empirical basis by comparing Rorschach scores of improved and unimproved schizophrenics. There is a striking paucity of hypotheses in these studies. Since most of the patient samples are small and almost no cross validation is reported, it is obvious that, as Cronbach [5] has recently emphasized, the reliability of the results is open to question. Finally, in view of the fact that the groups differ from one another in age, sex, intelligence, subtype classification, etc., it is perhaps not surprising that different prognostic signs were obtained.

The hypotheses for the present study were derived mainly from the prognostic investigations of schizophrenics by psychiatrists. Among the favorable prognostic signs most commonly reported in the literature [3] were a relatively well-adjusted prepsychotic personality, and the presence of affective features in the clinical picture. These observations form the principal basis for our operational hypotheses.

Problem

The aim of the present study was to investigate the value of the Rorschach in predicting the outcome of treatment of paranoid schizophrenics. It was assumed that the conflicting results obtained in previous prognostic investigations were due, in part, to the inadequate control of various factors. Among the most important of these factors, it was believed, is that of subtype classification. In the present study only patients diagnosed as paranoid schizophrenics were included.

This investigation differed from most of the other prognostic studies in at least one significant respect. No attempt was made to arrive at a combination of signs to distinguish improved from unimproved groups on a purely empirical basis. The purpose of all of the group comparisons made here was to test certain hypotheses or predictions, and at the same

---

[1] Reviewed in the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the author are the result of his own study and do not necessarily reflect the opinion or policy of the Veterans Administration.

The writer is indebted to the following for their assistance and encouragement: Dr. Donald W. Fiske and Dr. James G. Miller, University of Chicago, and Dr. Roy Brener, Veterans Administration Hospital, Hines, Ill.

time to re-examine certain prognostic signs cited by previous investigators.

The main reliance in this study was on relatively "objective" signs and techniques of measuring Rorschach variables. The author is aware of objections to this procedure that may be raised by clinicians, and of the limited validity and reliability of the "scales" used. Nevertheless, it is felt that such methods are most feasible in the present state of knowledge, and in view of the similar techniques applied in previous research in this area.

One chief hypothesis was that schizophrenics who improve should reveal evidence of a greater degree of anxiety in their Rorschach records than those who fail to improve. This hypothesis was based on an inference from one of the most frequently noted clinical observations, that schizophrenics who improve are more likely to manifest affective reactions (presumably including anxiety) than those who fail to improve. Furthermore, it seems reasonable that anxiety in a schizophrenic is indicative of an active struggle with the psychotic process. Such an attempt to regain contact with reality is in sharpest contrast to the apathy and withdrawal exhibited by typical deteriorated schizophrenics.

Another major hypothesis was that schizophrenics who improve should give evidence of having possessed a relatively better adjusted prepsychotic personality than unimproved patients. This would conform to the conclusion of Klopfer and Kelley regarding prognostic signs in the Rorschach: "The more the pretreatment record approaches 'normal,' the better the prognosis" [14, p. 368]. It is also in agreement with results of investigations of prepsychotic histories of schizophrenics.

A minor hypothesis considered was that improved patients should show less evidence of organic brain damage than unimproved. This supposition is based on the belief held by Bellak [4] and others that schizophrenia may be either of organic or psychogenic origin, or may involve varying degrees of each factor, and that the more favorable prognosis would be expected in those cases in which there are fewer organic factors.

Method

To test the hypothesis that anxiety is a favorable prognostic factor, Elizur's Rorschach content test [7] was employed. This test was designed to measure degree of anxiety and hostility manifested by the content of Rorschach responses. In view of the reliability and validity reported by Elizur, this device was considered appropriate.

Since it was believed that measures of "neurotic tendency" as indicated by Rorschach responses would give another indication of anxiety, the "neurotic signs" of Miale and Harrower-Erickson [10] were also used in this study.

To test the prediction that improved schizophrenics would show more nearly "normal" responses in their pretreatment Rorschach than unimproved patients, the following Rorschach factors were compared: total number of responses, distribution of whole and detail responses, number of human movement, color, shading, good form, space, human, and anatomy responses. As these components have been those most frequently compared in past studies, it is possible to relate the results to those of other investigators.

In addition, the data were analyzed for the most recent Rorschach prognostic signs reported by Phillips [17] and Piotrowski [21]. Records were scored for Phillips' "Color Drive," a scale purporting to measure "affecttive drive" based on the subjects' color responses. The Rorschach protocols were also scored for the six Piotrowski signs: "Variety," "Generic Term," "Evidence," "Color Response," "Indirect Color Approach," and "Demurring." These terms are defined and illustrated in the last of Piotrowski's reports [21].

Source of data and subjects. All clinical records of patients diagnosed as paranoid schizophrenics who had pretreatment Rorchach tests during 1947 and 1948 and were treated at Hines VA Hospital were examined. In selecting the records for the present study, the following criteria were used rigorously:

1. All patients included had a final diagnosis of "paranoid schizophrenia" after at least three months of hospitalization. This diagnosis was made by two psychiatrists, one of whom was a consultant. (In the Veterans Administration Hospitals a "consultant" is a specialist who is affiliated with a university and has attained a recognized standing in his field.)

2. There was positive evidence in the case record that the patient had hallucinations or bizarre delusions before treatment.
3. Patients were classified as "improved" if they were discharged from the hospital as in remission or improved and gave evidence that they no longer experienced delusions or hallucinations. The "unimproved" group consisted only of patients who were recommended for transfer to an institution for long-term hospitalization.

In addition to meeting the criteria described, each patient in the improved group was matched in age with a patient in the unimproved group. Out of 140 Rorschach records of paranoid schizophrenics, 36 patients were found, 18 improved and 18 unimproved, who fulfilled all these requirements. All the patients are males and veterans of World War II; all received some form form of shock treatment (insulin or electroshock) during the period from 1947 to 1949.

The two groups were similar not only in age but also in IQ and in years of schooling. Mean ages for the two groups were 28.1 and 28.3; mean IQ's, 98.8 and 101.4; mean years of schooling, 9.2 and 10.7. No statistically significant differences in intelligence or in education were found. In both groups a small number had histories of previous hospitalization for mental illness: seven of the improved and four of the unimproved. In addition to shock treatment, four patients in each group were also given some form of psychotherapy. The only important difference between the two groups was in duration of hospital treatment, the unimproved group remaining almost twice as long as the improved group. This is related to the fact that more of the unimproved than of the improved patients received both insulin and electroshock therapy. Patients who fail to benefit from one form of shock therapy are often given another form.

The Rorschach examinations were administered and scored or supervised by experienced clinical psychologists. In each case the examinations were given prior to treatment. The protocols were typed and all scoring and identifying information of both patient and original examiner were removed. The records were then rescored independently by psychologists other than the Rorschach administrator. There were thus two independent scores for every Rorschach response. The present writer checked each record in detail and, where differences in scoring occurred, made the final decision. The rescoring was done, of course, without knowledge as to whether the record belonged to the improved or unimproved patients. The records were scored in accordance with Beck's [3] system. Animal movement responses and form percentages as defined by Klopfer [14] were added for special scoring purposes.

Results

Rorschach scores as prognostic indicators. Table 1 presents the mean pretreatment scores of improved and unimproved patients on various Rorschach components. The only significant differences are as follows: The improved patients produced fewer total responses, higher *W*%, and fewer *D*'s than did the un-

Table 1

Mean Scores of 18 Improved and 18 Unimproved Paranoid Schizophrenics on Rorschach Components

Rorschach components	Mean of improved	Mean of unimproved	CR ("sign test")†	*t* (Mean Difference)
R	17.3	29.1	2.12	2.60*
W	6.6	5.7	0.71	0.78
W %	42.6	25.6	2.59*	2.60*
D	9.6	17.3	2.36*	3.23**
D %	51.6	60.9	1.89	1.37
Dd	1.1	6.1	1.65	2.00
Dd %	5.9	13.6	1.42	1.97
S	0.9	2.9	0.71	1.92
M	1.3	2.2	1.42	1.78
FY	1.2	1.8	1.65	1.52
Sum Y	1.3	2.3	0.24	0.92
FC	0.5	1.0	1.89	1.47
CF	1.1	1.8	1.18	1.22
Sum C	2.1	3.1	0.47	1.26
F plus %	77.1	70.9	0.94	1.39
H	1.3	2.4	0.94	1.59
H %	12.1	17.9	0.47	1.24
An	1.7	2.0	0.71	0.41
An %	8.9	6.8	0.24	0.63

* Significant at the .05 level of confidence.

** Significant at the .01 level of confidence.

† The "sign test," described by Johnson [12] does not make any assumption regarding normality of distribution (as does the *t* test). Since many of the Rorschach variables may not be normally distributed, it was decided to use both tests of significance. These sigma values are approximations, based on the "normal curve." Since the data used in computing signs are not continuous, significance was established by referring to the table of Dixon and Massey [6].

improved group. The other differences were statistically insignificant. The trend of the remaining differences may be summarized as follows: Improved patients exhibited fewer human movement, fewer shading, and a smaller number of color responses than unimproved patients. Since the latter factors depend to some extent on total number of responses, these results were not entirely unexpected. The smaller number of total responses given by improved patients, together with the trend toward a restriction of color and movement responses, suggested that Rorschach records of improved patients tend to resemble the records of neurotic more than they do those of normal individuals. The prediction that improved patients would produce more nearly "normal" records than unimproved patients is thus not borne out by the data. This finding is opposed to the evidence from many previous investigations and to the previously cited conclusion of Klopfer and Kelley [14].

"Neurotic signs" as prognostic indicators. In order further to test the hypothesis that improved patients would demonstrate more evidence of neurotic tendency than unimproved patients, the revised check list of Harrower-Erickson [10] was applied to the Rorschach data. Improved patients scored higher on the Harrower-Erickson check list of "neurotic signs" than unimproved patients: improved patients' score, 6.97; unimproved patients' score, 5.64. The difference between the two groups, however, was not statistically significant. The general trend of the data was, nevertheless, in the direction of confirming the hypothesis, and consistent with the restricted Rorschach records of the improved patients and with the results to be described in the next section.

Anxiety and hostility as prognostic signs. One of the principal hypotheses was, it will be recalled, that improved patients would show greater anxiety than unimproved patients. Elizur's [7] Rorschach content test was selected to measure anxiety. Since this test also is designed to measure degree of hostility on the basis of content, the latter variable was also scored for both groups on an exploratory basis.

Since the scores on the RCT are calculated by adding the scores of separate Rorschach responses, it can be seen that there is a direct relationship between total number of responses and total anxiety or hostility scores. In his report Elizur indicated that only records containing a specified range of scores were used. In view of the disparity in total number of responses between the two groups, it was necessary to make a correction by equalizing the total response scores, by totaling the separate scores and dividing by the total number of responses for each Rorschach record. The results were then expressed in terms of average anxiety or hostility score per response.

The results were as follows: The mean anxiety score per response for the improved patients was .44; for the unimproved, .23. The t ratio of 4.02 is significant at the .01 level of confidence. According to the "Sign Test," (See Note, Table 1), the significance also reaches the .01 level. In 15 of 18 comparisons, improved patients scored higher on the anxiety scale than unimproved patients.

The mean hostility score per response for improved patients was .07; for the unimproved, .17. The t ratio was 2.00, which approaches the .05 level of confidence. On the basis of the "Sign Test" it is significant at the .05 level. The improved patients showed less hostility than the unimproved group on this scale.

The prediction that improved patients would exhibit more anxiety in their Rorschach records than unimproved patients was, therefore, fulfilled. An additional finding was that improved patients also show a definite trend toward a lower degree of hostility than unimproved patients.

Prognostic value of "Organic signs." One hypothesis suggested by the literature is that schizophrenics who fail to recover should show more evidence of brain damage than those who do. To test this hypothesis, the data were evaluated in terms of Piotrowski's [18] "Organic signs." The results failed to confirm this hypothesis. The mean number of organic signs for the improved was 3.17; for the unimproved, 2.67. The difference is not statistically significant.

Prognostic signs found by Phillips and Piotrowski. An attempt was made to analyze the Rorschach data of our paranoid schizophrenics on the basis of Piotrowski's [21] six prognostic signs (Variety, Generic Term, Evidence, Color Response, Indirect Color Approach, and Demurring). In spite of Piotrowski's detailed description of these signs, the present writer experienced considerable difficulty in making judgments regarding their existence in many cases. Nevertheless, it is felt that, on the whole, the results are fairly reliable.

The attempt to analyze our Rorschach data by means of Piotrowski's six signs yielded negative results. The improved patients' records contained 4.3 signs; the unimproved, 4.1 signs. All but one patient in each group attained three or more signs, in contrast to Piotrowski's finding that most of his improved schizophrenics gave three or more signs, while very few of his unimproved patients did so. Furthermore, none of the six signs distinguished the improved from the unimproved patients at a statistically significant level.

Similarly negative results were obtained when Phillips' three signs were applied to the Rorschach data. These three signs were: "Demurring" (one of Piotrowski's six signs), *W%* between 15 and 25, and "Color Drive" score. There was no difference in the *W%* criterion, four patients in each group responding with *W*'s between 15 and 25%. The total "Color Drive" score for the improved group was 2.8; for the unimproved, 2.7. Scores per response (correcting for disparity in total number of responses) likewise showed no difference between the groups: mean score per response for the improved group being .10; for the unimproved, .09.

Thus, neither the prognostic signs of Phillips nor those of Piotrowski successfully distinguished the improved from the unimproved paranoid schizophrenics.

Discussion

The finding that evidence of anxiety in the Rorschach records of paranoid schizophrenics is a favorable sign is in keeping with the clinical observations of psychiatrists. As already mentioned, the presence of affective factors in the clinical picture is one of the most frequently noted characteristics of schizophrenics who later recover [4]. Since anxiety would be considered an affective reaction, the results confirm clinical-psychiatric data on prognostic indicators. Perhaps anxiety is a favorable indicator because it implies a struggle with the psychotic process. Such a condition contrasts markedly with the inertia, blunting of affect, and general apathy characteristic of chronic and deteriorated schizophrenics.

Previous Rorschach prognostic studies rarely mention anxiety as a prognostic sign. Nevertheless, some of the signs found by Piotrowski and Phillips, though they are not designated as such by these authors, could be interpreted as indicative of anxiety. Piotrowski's [21] description of two of his signs, "Evidence" and "Demurring," suggests that basically these factors may be an expression of anxiety. "Evidence" is scored when the subject indicates he is weighing the evidence regarding adequacy of response by using terms expressing doubt. "Demurring" relates to the subject's holding back responses in order not to commit himself to interpretations of which he is not certain.

Similarly, many features of Phillips' [17] "Color Drive Scale" are identical with Elizur's criteria for scoring anxiety in the Rorschach. Both scales give positive weights to responses like "blood," "fire," and others suggesting stress and danger.

The other positive findings—that hostility is an unfavorable sign—has also not specifically been mentioned in the literature of Rorschach investigators. Two recent clinical studies of delusions of schizophrenics do, however, tend to substantiate this result. Seitz [23] and Albee [1], in separate investigations, reported that there is greater likelihood for improvement if a schizophrenic's delusions manifest more intrapunitive than extrapunitive mechanisms. This finding would at the same time corroborate the previous conclusion that anxiety is prognostically favorable.

The second major hypothesis—that Rorschach signs pointing to a relatively good prepsychotic adjustment indicated a favorable outcome—was not confirmed by the data. This lack of confirmation of previous Rorschach findings requires some explanation. It should be noted that, while there is considerable variation in specific signs cited as prognostic by different investigators, there seems to be a definite tendency for "favorable" Rorschach signs to be associated with a good prognosis in the majority of the investigations. Why should the paranoid group show a reversal of this trend?

One possible explanation lies in the nature of the main positive finding, that anxiety is a favorable sign. As might be expected, anxiety resulted in a constricted Rorschach record for most of the improved patients. Such a record would show lowered productivity and consequently a reduction or elimination of those movement or color responses that reveal the more positive aspects of personality. In other words, the effect of anxiety would be to inhibit the expression of favorable personality factors. Perhaps the paranoid group, being the subtype in which control is the most rigid [2], is the one where the presence of anxiety, if it exists, would be the most conspicuous favorable sign. In other varieties of schizophrenia, where there is less intellectual control, there may be a greater opportunity for release of those responses in the Rorschach which are indicative of inner resources and capacity for emotional rapport.

In view of the ill-defined nature of the diagnostic classification of schizophrenia, there is a great need for precise descriptions of the particular group selected for observation, as well as of the type of treatment given. Better measures of results of treatment are also required. Although the present data fall far short of ideal in all these respects, the investigation was limited to a fairly well-defined subgroup of schizophrenia with specified behavioral characteristics. Confirmation of the results, which is necessary if reliance is to be placed on them, should come from a repetition of this type of investigation on a similar group of paranoid schizophrenics.

Summary

This investigation dealt with the prognostic value of the Rorschach in paranoid schizophrenia. It was felt that previous prognostic studies were inadequate because of failure to control various factors, particularly that of subtype classification. The chief aim of this investigation was to test certain hypotheses regarding prognosis from Rorschach records of paranoid schizophrenics.

The pretreatment Rorschach records of 36 diagnosed paranoid schizophrenics, who were subsequently given shock treatment, were analyzed to determine the presence of prognostic signs that would distinguish the improved from the unimproved patients. The sample consisted of 18 improved and 18 unimproved subjects matched, by pairs, for age. There was no statistically significant difference between the two groups in IQ and in educational level.

The following results were obtained: The pretreatment Rorschach records of improved paranoid schizophrenics contained evidence of a greater degree of anxiety than did the records of unimproved patients. Mean anxiety score per response on Elizur's Rorschach content test differentiated the two groups at a statistically significant level. Improved patients also manifested significantly less hostility than unimproved patients. There was some tendency, though not statistically significant, for improved patients to exhibit more "neurotic signs" in their Rorschach records than unimproved patients. On the average, Rorschach records of improved paranoid schizophrenics tended to be less productive and to contain fewer movement and color responses than those of the unimproved group. The latter results are in harmony with the previously cited findings of greater anxiety in the improved patients.

No difference was found between improved and unimproved patients in total number of Piotrowski's "organic signs" in their Rorschach records. An effort to confirm specific Rorschach prognostic signs found by previous investigators yielded negative results.

An attempt was made to rationalize the positive findings in this study and to integrate them with previous prognostic indicators reported by psychiatrists and clinical psychologists.

Received November 3, 1952.

References

1. Albee, G. W. The prognostic importance of delusions in schizophrenia. *J. abnorm. soc. Psychol.*, 1951, **46**, 208–212.
2. Alexander, F. *Fundamentals of psychoanalysis.* New York: Norton, 1948.
3. Beck, S. J. *Rorschach's test.* Vol. I. New York: Grune & Stratton, 1946.
4. Bellak, L. *Dementia praecox.* New York: Grune & Stratton, 1948.
5. Cronbach, L. J. Statistical methods applied to Rorschach scores: A review. *Psychol. Bull.*, 1949, **46**, 393–429.
6. Dixon, W. J., & Massey, F. J. *An introduction*

to statistical analysis. New York: McGraw-Hill, 1951.

7. Elizur, A. Content analysis of the Rorschach with regard to anxiety and hostility. *Rorschach Res. Exch.*, 1949, **13**, 247–284.
8. Graham, Virginia T. Psychological studies of hypoglycemia therapy. *J. Psychol.*, 1940, **10**, 327–358.
9. Halpern, Florence C. Rorschach interpretation of the personality structure of schizophrenics who benefit from insulin therapy. *Psychiat. Quart.*, 1940, **14**, 826–833.
10. Harrower-Erickson, Molly R. The value and limitations of the so-called "neurotic signs." *Rorschach Res. Exch.*, 1942, **6**, 109–114.
11. Henderson, D. K., & Gillespie, R. D. *A textbook of psychiatry.* New York: Oxford Univer. Press, 1944.
12. Johnson, P. O. *Statistical methods in research.* New York: Prentice-Hall, 1949.
13. Kisker, G. W. A projective approach to personality patterns during insulin-shock and metrazol convulsive therapy. *J. abnorm. soc. Psychol.*, 1943, **37**, 120–124.
14. Klopfer, B., & Kelley, D. McG. *The Rorschach technique.* Yonkers, N.Y.: World Book Co., 1942.
15. Morris, W. W. Prognostic possibilities of the Rorschach method in metrazol therapy. *Amer. J. Psychiat.*, 1943, **100**, 222–230.
16. Pacella, B. L., Piotrowski, Z., & Lewis, N. D. C. The effects of electric convulsive therapy on certain personality traits in psychiatric patients. *Amer. J. Psychiat.*, 1947, **104**, 83–91.
17. Phillips, L. Personality factors and prognosis in schizophrenia. Unpublished doctor's dissertation, Univer. of Chicago, 1949.
18. Piotrowski, Z. The Rorschach inkblot method in organic disturbances of the central nervous system. *J. nerv. ment. Dis.*, 1937, **86**, 525–537.
19. Piotrowski, Z. Prognostic possibilities of the Rorschach method in insulin-treated schizophrenics. *Psychiat. Quart.*, 1938, **12**, 679–689.
20. Piotrowski, Z. A simple experimental device for the prediction of outcome of insulin treatment in schizophrenia. *Psychiat. Quart.*, 1940, **14**, 267–273.
21. Piotrowski, Z. The Rorschach method as a prognostic aid in the insulin-shock treatment of schizophrenics. *Psychiat. Quart.*, 1941, **15**, 807–822.
22. Roe, Anne, & Shakow, D. Intelligence in mental disorder. *Ann. N.Y. Acad. Sci.*, 1942, **42**, 361–490. Also in S. S. Tomkins (Ed.), *Contemporary psychopathology.* Cambridge: Harvard Univer. Press, 1943. Pp. 348–354.
23. Seitz, P. F. D. A dynamic factor correlated with the prognosis in paranoid schizophrenia, *Arch. Neurol. Psychiat., Chicago,* 1951, **65**, 604–606.
24. Wittman, Phyllis. The use of the multiple choice Rorschach as a differential diagnostic tool. *J. clin. Psychol.* 1945, **1**, 281–287.
25. Windle, C. Psychological tests in psychopathological prognosis. *Psychol. Bull.*, 1952, **49**, 451–482.

Journal of Consulting Psychology
Vol. 17, No. 3, 1953

A Comparison of Scores on the Index of Adjustment and Values with Behavior in Level-of-Aspiration Tasks[1]

Robert E. Bills
University of Kentucky

Many writers have expressed the opinion that level of aspiration, obtained from experimental studies, samples a type of behavior which is characteristic of subjects' goal-setting behavior in general. In the experimental tasks, a subject is given practice and, after being informed of his level of performance, is asked to state his level of aspiration for the next trial. This is the most common method for assessing level of aspiration. Another possible method of measuring level of aspiration is by determining the discrepancy which exists between the concept of self and the concept of the ideal self. The latter type of aspiration is a resultant of the self percepts and the ideals held by a person, and may or may not be comparable with the level of aspiration represented by the simple tasks usually used in level-of-aspiration studies. One of the purposes of the present study was to investigate the extent of the relationship between level of aspiration as determined by motor and verbal tasks and personal level of aspiration as determined by the Index of Adjustment and Values [3].

The Index of Adjustment and Values. The Index of Adjustment and Values requires that a subject make three ratings, on a five-point scale, for each of 49 traits. These ratings have been grouped into three measures: concept of self, acceptance of self, and concept of the ideal self. The discrepancy between concept of self and concept of the ideal self is a measure of the level of aspiration set by a subject in respect to his ideals. The Index has been shown to be valid as a measure of emotionality [2, 9], and is capable of distinguishing personality types [1].

Hypotheses of the study. It was predicted, on the basis of the reasoning described above, that personal level of aspiration as revealed by the Index would be significantly correlated with level of aspiration as determined by experimental tasks of a motor and verbal character.

Eysenck [4], in a study of the level-of-aspiration behavior of hysterics and dysthymics, demonstrated that his hysterical cases were more variable in the levels of aspiration they set than were his dysthymic cases.

A Rorschach study [1] of high and low scorers on acceptance of self as measured by the Index of Adjustment and Values showed that high and low scorers constituted two distinct personality groups, and that the low scorers showed more rigid and less variable Rorschach records than did the high scorers. On the basis of this result and Eysenck's findings it was predicted that a group of subjects with acceptance-of-self scores below the population mean of the Index would be less variable in the level of aspiration they set than would a group characterized by acceptance-of-self scores above the population mean of the Index.

It should be possible in a controlled setting, such as a level-of-aspiration experiment, to demonstrate relationships between personality measures and observations of behavior. In a previous study [3] it was shown that acceptance of self is related to direction of perceived threat. Subjects who scored above the population mean of acceptance of self as measured by the Index of Adjustment and Values blamed factors outside themselves for unhappiness they had experienced, but subjects below the mean blamed themselves. It was predicted, therefore, in the present study that subjects who scored above the mean on acceptance of self, when questioned about their performance, would respond differently from subjects who were below the mean on acceptance of self. More specifically, it was hypothesized that: (*a*) Subjects who score below the population mean on acceptance of self, as shown by the Index, will underestimate their performances more often than subjects who score above the mean on acceptance of self; (*b*) subjects who score above the population mean on acceptance of self, when asked to recall their level of performance after an interpolated task, will remember their performance as being above their true performance level and subjects who score below the mean will recall their level of performance as being lower than it was; (*c*) subjects who are below the population mean on acceptance of self, when asked, "How do you feel about your performance?" will express negative, internalized attitudes while subjects who are above the population mean will express positive, externalized attitudes.

---

[1] This study was made possible through a grant by the University of Kentucky Research Fund. The author wishes to express his appreciation to the committee in charge of the fund and to Glen E. Roberts who worked as assistant on the project.

Design

Thirty female subjects were selected, from among students at the University of Kentucky, on the basis of discrepancy scores on the Index of Adjustment and Values. The selected subjects had scores distributed from high to low on discrepancy.

Each subject was given, individually and in one session, five level-of-aspiration tasks with the order of presentation of the tasks randomized to prevent serial effects. The tasks were selected to represent a variety of motor and verbal activities and mixtures of the two. The five tasks included: *1,* dart throwing; *2,* Rotter target aspiration board [11] as modified by Lonstein [8]; *3,* marking out letters; *4,* substituting letters; and *5,* addition.

On the dart-throwing task each subject was given five darts, and the score was the number of points scored on all five darts. The modified Rotter board was a board with a groove to accommodate a marble, on which depressions were made at regular intervals to stop the marble. The center depression scored six and depressions on either side were on a descending scale to one point. Each subject rolled the marble five times for each trial. In the marking-out test, the subject was given a paragraph and instructed, "Mark out as many *e's* as you can. You have two minutes." The substitution test consisted of a series of letters with a blank space beneath each. At the top of the page was a key. Subjects were instructed to place below each letter the correct letter according to the key. Two minutes were allowed on each series. The arithmetic tests were series of five two-place numbers. The digits were arranged in a different order in each series but the problems on all the series contained the same combinations. The time limit was two minutes. At the beginning of each task the subject was requested not to attempt to add his own score.

In each task the subject was given a series of practice trials and then informed of his score on the last practice trial. He was then told, "The next trial is a test. How much do you expect to make?" The subject was told his score on the test trial and his performance was recorded. He was then asked, "How much do you think you should make the next time?" Performance was recorded and the subject was asked, "How much do you think you scored?" The subject was then told, "You scored. . . . This is your score on the . . . test. How do you feel about your performance?" The subject's remarks were recorded verbatim.

After the five tasks had been completed, the subject was handed a mimeographed page and instructed, "Read this as rapidly and carefully as possible." The experimenter pretended to make notes and to time the performance. This exercise was included to separate the tasks from recall of performance. After the subject had completed the reading of the excerpt, he was asked to recall his performance on each task. Recall was recorded and certainty of recall was judged on a five-point scale by the experimenter.

Results

The experiment yielded data of four types including: level-of-aspiration scores, estimates of performance, recall of performance, and attitude toward performance. These will be discussed below.

Index scores and level of aspiration. Subjects were ranked on the two Index scores and according to level of aspiration shown by each of the five tasks. Level of aspiration was defined operationally to be the difference between the first test performance and the goal set by the subject for his second test performance. Rank-order correlation coefficients for acceptance-of-self scores and discrepancy scores of the Index of Adjustment and Values and the five measures of level of aspiration are given in Table 1. The data in Table 1 show that the two scores of the Index are not highly correlated with level of aspiration in the five measures used. Only two of the correlations were significantly different from zero at the .05 level or less. The correlation of discrepancy score and mark-out test was significantly different from zero at the .05 level of confidence and the correlation of acceptance-of-self scores and Rotter board scores was significantly different from zero at the .01 level.

Table 1

The Rank-Order Correlation Coefficients between the Index Scores and Level of Aspiration Shown by the Experimental Tasks

Tasks	Acceptance of Self	Discrepancy
Darts	.18	–.24
Rotter	.51**	–.15
Addition	.08	–.07
Substitution	.13	.04
Mark-out	.01	.36*

* Significantly different from zero at the .05 level.
** Significantly different from zero at the .01 level.

Intercorrelations of level-of-aspiration tasks. In order to compare levels of aspiration as shown by each of the five experimental tasks the rank orders for all of the tasks were intercorrelated. The intercorrelations are low but with the exception of the dart task, which did not correlate significantly with any of the other tasks, are significantly different from zero at the .05 level or less. The intercorrelations of the level-of-aspiration tasks appear to be higher than the correlations of the Index scores and the level-of-aspiration scores. The intercorrelations of the level-of-aspiration tasks are contained in Table 2.

Table 2

Rank-Order Intercorrelations among the Five Experimental Tasks

	Rotter	Addition	Substitution	Mark-out
Darts	–.01	–.08	.02	.30
Rotter		.39*	.39*	.44*
Addition			.49**	.43*
Substitution				.39*

* Significantly different from zero at the .05 level.
** Significantly different from zero at the .01 level.

Index scores and variability of level of aspiration. The distributions of levels of aspiration in the five tasks were converted into standard scores so that scores on the tasks would be comparable. This manipulation resulted in five distributions of standard scores, one for each task. Each of these five distributions was divided into two groups dependent upon the subject's score (above or below the mean) on acceptance of self on the Index of Adjustment and Values, and into two other groups according to discrepancy scores above and below the population mean on the Index. Variances of each of the pairs of distributions were compared by means of the F test. These tests show that on two of the five tasks, subjects with acceptance-of-self scores above the population mean of the Index were significantly more variable in the levels of aspiration they expressed than were subjects who had acceptance-of-self scores below the mean. Likewise, subjects with discrepancy scores below the mean were more variable than subjects with

Table 3

Variabilities of Levels of Aspiration when Subjects Are Divided According to Acceptance-of-Self Scores on the Index of Adjustment and Values

	Variance			
Task	Acceptance of Self Above Mean* $N=13$	Acceptance of Self Below Mean $N=17$	F	p
Darts	1.188	1.414	1.190	N.S.
Rotter	.913	1.036	1.135	N.S.
Mark-out	1.612	.653	2.469	.05
Substitution	1.419	.196	7.240	.001
Addition	.822	1.190	1.448	N.S.

* Acceptance-of-self scores on the Index of Adjustment and Values above the mean of the standardization population of the Index.

scores above the mean. Acceptance-of-self scores on the Index and discrepancy scores, when correlated, yield coefficients around −.75. These data are presented in Tables 3 and 4.

Table 4
Variabilities of Levels of Aspiration when Subjects Are Divided According to Discrepancy Scores on the Index of Adjustment and Values

Task	Variance Discrepancy Below Mean* $N=16$	 Discrepancy Above Mean $N=14$	F	p
Darts	1.185	1.400	1.181	N.S.
Rotter	1.722	.331	5.202	.01
Mark-out	1.068	.914	1.168	N.S.
Substitution	.478	1.061	2.220	N.S.
Addition	1.528	.538	2.840	.05

* Discrepancy scores on the Index of Adjustment and Values below the mean of the standardization population of the Index.

Acceptance of self and estimate of performance. Subjects were asked to estimate their performance on each of the tasks as soon as they had completed each task. A dichotomous chi square was calculated using the mean acceptance-of-self score of the standardization population of the Index to form one dichotomy and the number of tasks in which performance was underestimated as the other dichotomy. The latter dichotomy was formed by dividing the number of underestimates of all the subjects at the median number of underestimates. This placed all subjects with four or five underestimated performances in one category and subjects with three, two, or one underestimated performance in another category. All subjects underestimated at least one performance. Two subjects who were above the mean on acceptance of self underestimated four or five performances as compared with ten subjects below the mean, while ten subjects who were above the mean on acceptance of self underestimated three, two, or one performance as compared with eight subjects below the mean. This combination yielded a chi square of 4.535 which is significantly different from zero at the .05 level, and may be interpreted to mean that estimate of performance is related to acceptance of self.

Acceptance of self and recall of performance. At the end of the experiment, following the reading of the excerpted paragraphs, each subject was asked to recall his scores on each of the tasks. The number of overestimates, correct estimates, and underestimates was calculated for each subject and the subjects were divided into "overestimators" and "underestimators." The one subject who had more correct than faulty recalls was eliminated from consideration. These two groups formed one dichotomy of a double dichotomy. The other division was formed by dividing the subjects' acceptance-of-self scores at the mean acceptance-of-self score of the standardization population of the Index.

Eight subjects with acceptance-of-self scores above the mean were classed as "overestimators" while only four subjects below the mean were included in this category. Four subjects with acceptance-of-self scores above the mean were included as "underestimators" but 13 subjects with scores below the mean were grouped in this division. This combination produced a chi square of 5.396 which is interpreted at the .05 level to show that recall of performance is related to acceptance of self.

Acceptance of self and certainty of recall. Each subject was rated on the degree of certainty for each of his five recalls. Subjects were divided according to acceptance-of-self scores above and below the standardization mean and certainty of recall using the median of the averages for the 30 subjects as the dividing point. A double dichotomous chi square was calculated but was not significant.

Acceptance of self and attitude toward performance. Each subject was asked to state how he felt about his performance in each of the tasks. Lists of the verbatim comments of the subjects were given to five judges who were asked to classify each statement.[2] The judges decided whether the attitude expressed in the statement was positive or negative and whether it was directed toward or away from self. In this way five judges rated five statements for each subject. The subjects were then categorized according to the classification which appeared most frequently as negative-internal, negative-external, positive-internal, and positive-external. A double dichotomous chi square was calculated using negative and positive attitudes as one dichotomy and scores above and below the population mean on acceptance of self as the other dichotomy. One subject with an acceptance-of-self score above the mean expressed a negative attitude toward his performance compared with 12 subjects below the mean, but 12 subjects with acceptance-of-self scores above the mean demonstrated positive attitudes while only 5 subjects below the mean showed positive attitudes. The chi square produced by this combination was 11.868, which is significant at less than the .01 level. Thus, type of attitude expressed toward performance in the level-of-aspiration tasks is related to acceptance of self.

---

[2] The author wishes to express his thanks to Virginia Parko, Thomas Sutherland, Travis Rawlings, and Glen E. Roberts who served as judges.

A chi square was calculated using acceptance of self as one dichotomy and direction of expressed attitude as the other dichotomy. Eight subjects with acceptance-of-self scores above the population mean directed their attitudes about their performance toward themselves and 5 directed it away from themselves, while 16 subjects below the mean directed their attitudes toward themselves and 1 directed it away from himself. This chi square was equal to 4.887, which is significant at the .05 level, and shows that acceptance of self is related to the direction of expressed attitude.

Discussion

The intercorrelations of the Index and the experimental tasks. Eysenck has stated:

> The actual size of the correlation [between two levels-of-aspiration tasks] would appear to depend on three points:
>
> (1) The similarity of the two tests, as regards the ability measured.
>
> (2) Similarity of the scale on which goodness of performance is measured.
>
> (3) Similarity of the experimental situation, i.e., whether the two tests are given in the same experimental session or not [4, p. 133].

Studies with normal subjects have reported correlations between level of aspiration on different tests ranging from .29 by Gould [7] to .70 by Frank [5] with an estimated median of .40 to .45.

The intercorrelations of the five tasks included in the present study approximate the average of the reports in the literature and are about as high as might be expected when examined in the light of Eysenck's statement. The tasks were purposely selected to represent a variety of motor and verbal performances, and only the last one of Eysenck's three generalizations was applicable to this experimental situation, namely, that all of the experimental tasks were given in one session.

The data show that the Index scores were not highly correlated with level of aspiration as measured by the five tasks, but the correlations approach the size of the intercorrelations among the five experimental tasks. Several possible explanations of the low degree of intercorrelation between the tasks, and the correlations of the Index scores with the tasks appear tenable: (*a*) The Index is not a valid measure of level of aspiration. (*b*) Ego involvement in a task is an important variable in determining the usefulness of a task as a measure of level of aspiration. (*c*) There may be more than one type of level of aspiration and each of the tasks and the Index may be measures of general and specific types of level of aspiration. (*d*) The level of aspiration as measured by the Index may be a more exact measure than the levels shown by the experimental tasks. These possibilities will be discussed below.

In regard to the first point, the Index of Adjustment and Values has not been validated as a measure of level of aspiration, but experimental tasks which are commonly used as measures of this variable have not been validated for this purpose either. It seems logical to assume, however, that the discrepancy score of the Index is as valid a measure of level of aspiration as the usual experimental tasks. The subject in filling out the Index is asked on 49 items, after he has made a rating of his present status, to tell how he would like to be. This process is equivalent to obtaining 49 estimates of his level of aspiration. This is essentially the same process by which level of aspiration is determined in experimental tasks.

On the second point, several authors [6, 11] have noted that the more a subject sees a level-of-aspiration task as a "game," the less likely will be a high correlation between this task and other measures. This may mean that the greater the extent of ego involvement in a task, the more valid the task will be as a measure of level of aspiration. The Index seems eminently fitted to the requirements of an ego-involved measure. All of the ratings on the Index are done by the subject, on himself, and from his own frame of reference.

It seems reasonable to assume that the personal level of aspiration of female college students would contain a large verbal factor, and thus, it becomes of greater interest that the highest relationship discovered between discrepancy scores and level of aspiration in the present study was shown in the "mark-out" test which is probably the most verbal task of the battery. The results indicate the possibility of the interaction of level of aspiration and concept of self. It is possible that the validity of level-of-aspiration measures is at least partially dependent upon the concept of self of the subject. Level of aspiration shown by a task is possibly more highly related to personal level of aspiration when performance in the task is seen by the subject as enhancing to self-organization. The importance of personal factors in determining level of aspiration has been demonstrated in numerous experiments.

The third and fourth points cannot be settled in this paper, but are probably worthy of further study involving factor analysis and validation with external criteria. "Would the Index or the usual type of level-of-aspiration tasks show higher agreement with external criteria?" is a question which might be investigated profitably.

The Index and variability of performance. Studies [1, 3] of the Index of Adjustment and Values have shown that the mean acceptance-of-self score of the standardization population separates two groups which are distinctly different in personality characteristics. Rorschach examination revealed that high scorers on acceptance of self have more variable personalities than low scorers, and the data from the present study provide additional evidence for this conclusion. In the two of the five tasks where the difference in the variability of the groups was statistically significant, the high scorers constituted a more variable group than the low scorers.

Acceptance-of-self scores on the Index are negatively and highly correlated with discrepancy scores [3]. When the variabilities of level-of-aspiration behavior of high and low scorers on discrepancy were compared, the low scorers were significantly more variable on two of the tasks than were the high scorers. These two tasks were not the same two which were significantly different when the division was made on the basis of acceptance of self, which may indicate that although the two Index scores are highly correlated, they are probably measures of different variables.

The Index and other observed behavior. The Index served as a better predictor of other behavior recorded in the experiment than it did for level of aspiration. Attitude toward self was shown to be related to attitude toward performance, direction of expressed attitude toward performance, estimate of performance, and recall of performance. These findings support the conclusion of other writers [10, 12] that attitude toward self is an important determinant of behavior.

Summary

Thirty volunteer, female subjects were tested with the Index of Adjustment and Values and five level-of-aspiration tasks. Subjects set levels of aspiration for each of the five tasks, estimated their performance in the tasks, expressed comments regarding their performance, and after a filled interval attempted to recall their performances.

It was concluded that Index scores were, to a low degree, related to level of aspiration, as measured by the experimental tasks, that the variability of the level of aspiration set by groups selected by the Index was significantly different, and that acceptance of self shown by the Index was significantly related to attitude toward performance, direction of expressed attitude toward performance, estimate of performance, and recall of performance.

Received July 23, 1952.

References

1. Bills, R. E. Rorschach characteristics of persons scoring high and low in acceptance of self. *J. consult. Psychol.*, 1953, **17**, 36–38.
2. Bills, R. E. A validation of changes in scores on the Index of Adjustment and Values as measures of changes in emotionality. *J. consult. Psychol.*, 1953, **17**, 135–138.

3. Bills, R. E., Vance, E. L., & McLean, O. S. An index of adjustment and values. *J. consult. Psychol.*, 1951, **15**, 257–261.
4. Eysenck, H. J. *Dimensions of personality.* London: Kegan Paul, 1947.
5. Frank, J. D. Individual differences in certain aspects of level of aspiration. *Amer. J. Psychol.*, 1935, **47**, 119–128.
6. Frank, J. D. The influence of the level of performance in one task on the level of aspiration in another. *J. exp. Psychol.*, 1935, **18**, 159–171.
7. Gould, R. An experimental analysis of "level of aspiration." *Genet. Psychol. Monogr.*, 1939, **21**, 1–116.
8. Lonstein, M. A comparative study of level of aspiration variables in neurotic, psychopathic, and normal subjects. Unpublished doctoral dissertation, Univer. of Kentucky, 1952.
9. Roberts, G. E. A study of the validity of the Index of Adjustment and Values. *J. consult. Psychol.*, 1952, **16**, 302–304.
10. Rogers, C. R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951.
11. Rotter, J. B. Level of aspiration as a method in the study of personality. I. A critical review of methodology. *Psychol. Rev.*, 1942, **49**, 463–474.
12. Snygg, D., & Combs, A. W. *Individual behavior.* New York: Harper, 1949.

Journal of Consulting Psychology
Vol. 17, No. 3, 1953

Directionality of Lines in the Bender-Gestalt Test

Roland M. Peek
Hastings State Hospital[1]

Although the use of the Bender-Gestalt test (B-G) at all age groups has increased in both clinical and educational settings, interpretational techniques with demonstrated validity have been glaringly lacking. Interpretation of B-G protocols has for the most part been highly subjective and consequently dependent upon the skill and range of experience of individual examiners whose administration and manner of interpretation may vary widely from that of other "experts" dealing with other populations. In most cases the communication of the personal-experience type of skill (assuming its validity) is hampered and distorted by the lack of objective and definitive terms with which to describe the behavior occurring in response to the stimulus figures.

One of the first needs in the study of the B-G, therefore, has been for some sort of scoring system which would clearly define each variable in objective terms so that the scoring would become satisfactorily reliable, therefore providing variables which were subject to experimentation; the scoring must also be simple enough as to be applicable to the needs of the time-pressed clinician. Any scoring system should be capable of this dual function. This means that the scoring could not be totally subjective as were pioneer systems [1, 3], nor yet so minutely objective that scoring time prohibited the use of the test [2]. Another difficulty which has occurred is that some systems function on a level which excludes much psychologically meaningful data, or combines variables which are not psychologically compatible, thus restricting the potential usefulness of the test both clinically and experimentally [5].

In the course of developing and validating a scoring technique aimed at a resolution of these problems [6], it was noted that a particular aspect of B-G behavior seemed to be of clinical significance, though not included in any scoring system previously published: the order and directionality used by the subject in drawing the parts of each figure. Such data can be included in a sheet containing general observations of the subject's comments, handedness, attitude, etc. during the testing.[2] It has seemed apparent that deviations from popular modes of drawing the figures have psychological significance independently of the shape of the resulting drawing, and that this information combined with other variables increases the range of predictable extra-test behavior and the accuracy of such predictions.

Problem

To test the hypothesis that variations from the usual approach reflect individual differences in personality, comparison was made between a control group and another group from the same population who drew the diagonal projection of Figure 5 in a specifically deviant way, i.e., starting at the top and drawing the dots in a series down to the half-circle, rather than extending the projection from the half-circle outwardly, as most subjects do.

Subjects

The control group consisted of 75 hospitalized male neuropsychiatric patients randomized by selecting every 20th discharged case in the files of a VA acute-treatment hospital. The 75 cases in the experimental group, who came from the same section of the same hospital, were selected solely on the basis of having been reported to the author by any of six psychological trainees as having drawn the projection on

---

[1] This study was made at the VA Hospital Annex, Ft. Snelling, St. Paul, Minnesota.

[2] The authors of the Peek-Quast scoring system for the B-G are greatly indebted to Dr. George S. Welsh for the original suggestion that these additional data may have potential clinical usefulness.

Figure 5 in the specified direction.[3] Since no investigation was made of the psychological testing of the controls, it is not known how many of that group drew Figure 5 as described.

Procedure

A check list was devised on which was tallied the information listed in Table 1. The source of the information was the brief final summary prepared by the physician when a patient was discharged from the hospital; because of its nature more comprehensive or more specific data were unfortunately not consistently enough noted to make worth while their inclusion in the check list. A given category was checked if one of the specific words in each category was used when referring to the behavior of the patient. The attributes were not tallied if described as deriving from the psychological testing without reference to overt behavior. The occurrences of each category were expressed in percentages; the statistical significance of the differences in percentages were determined by the use of phi coefficients as described by Jurgensen [4].

Table 1
Items from Discharge Summary
Included in Check List

1. Diagnosis
2. Treatment (EST, IST, Other)
3. Prognosis
    (*a*) good (*b*) fair (*c*) guarded (*d*) poor
4. Insight
    (*a*) good (*b*) fair (*c*) partial
    (*d*) poor (*e*) none
5. Number of somatic complaints
    (*a*) none (*b*) one (*c*) two
    (*d*) three (*e*) four (*f*) five
6. Number of somatic complaints (cumulative totals)
    (*a*) one or more (*b*) two or more
    (*c*) three or more (*d*) four or more (*e*) five or more
7. Headache
8. GI complaints
9. Cardiac complaints
10. Respiratory complaints
11. Musculature involvement
12. Weakness, dizziness
13. Fatigue
14. Anorexia
15. Weight loss (as presenting complaint, not merely as a physical finding)
16. Excessive perspiration (as a presenting complaint)
17. Dermatitis
18. Buzzing or ringing in ears
19. Other somatic complaints
20. Depressed
21. Suicidal thoughts or attempts
22. Blackouts; amnesia
23. Insomnia
24. Tremulous
25. Nervous, tense, restless, anxious
26. Tense (tallied separately)
27. Excessive drinking or addiction to drugs
28. Incapacitated or adjudged incompetent
29. Hallucinations or delusions
30. Impotence; homosexuality
31. Violent fears or dreams directed at patient (see text)
32. Paranoid
33. Seclusive
34. Immature, dependent, inadequate, inferiority feelings, passive
35. Impulsive, poor judgment
36. Hostile, aggressive
    (*a*) readily expressed
    (*b*) poorly expressed
37. Tearful
38. Divorce, extreme marital or sexual problem (if evident in overt behavior)
39. Parental deprivation or abnormality (death or separation of parents before age 12, or overt and severe problems such as being beaten regularly by a parent)
40. Number of tallies in Somatic Complaint categories plus items 34, 35, 36, 38
    (*a*) none (*b*) one (*c*) two (*d*) three or less (*e*) six or more (*f*) seven or more

Results

Table 2 lists those categories for which the difference in percentages reached statistical significance at the .001, .01, .05, and .10 levels of confidence, the latter included only to indicate the general trends running through the data.

No significant differences appear between the diagnoses of the two groups, neither in terms of specific diagnosis nor by broad psychotic-nonpsychotic categories. No differences are apparent in treatment received nor in the degree of insight attained. However, the control group shows a greater percentage whose prognosis is described as "good."

---

[3] Although subgroups were prepared on the basis of the direction of drawing the arc, and whether the projection was drawn first, the number of cases in each subgroup was too small at the time of the experiment to permit statistical analysis of the apparent differences between these subgroups.

Table 2
Items from Check List (see Table 1) Differentiating a Random Group of NP Patients from Patients Drawing Projection of B-G Fig. 5 toward the Arc
(*N* is 75 for each group)

Item	Controls %	Exp. %	Item	Controls %	Exp. %
Significant at .001 Level			Significant at.01 Level		
5(*a*)	32	5	6(*e*)	7	24
6(*b*)	45	83	7	16	37
6(*c*)	25	63	10	3	15
6(*d*)	19	44	11	31	52
20	31	63	15	1	12
34	25	55	22	4	19
36	25	52	31	0	11
37	3	21	35	11	31
41(*d*)	69	23	39	17	36
41(*e*)	5	37	40(*a*)	9	0
41(*f*)	3	23	40(b)	17	3
Significant at .05 Level			Significant at .10 Level		
3(*a*)	13	3	3(*a*)+3(*b*)	32	20
5(*d*)	7	19	3(*c*)+3(*d*)	68	80
5(*f*)	3	13	4(*a*)	11	4
6(*a*)	68	85	5(*b*)	23	12
14	9	21	5(*c*)	20	33
19	21	37	8	23	36
21	13	27	12	16	27
36(*a*)	23	41	26	37	51
40(*c*)	21	8	28	16	27
			36(*b*)	3	11

Striking differences are noted in the complaints listed in the final discharge summary. The control group is more likely to have no somatic complaints at all, while the experimental group begins to show a significantly greater number as the somatic complaints reach three and above. Using cumulative totals, the experimental group is more likely to have one or more, two or more, three or more, four or more, and five or more somatic complaints.

A difference appears not only in number but in the types of somatic complaints. A significantly greater proportion in the experimental group has headaches, respiratory ailments, complaints involving the musculature, severe weight loss, anorexia, and "other somatic complaints."

In other characteristics and complaints noted by the psychiatrists significant differences also occur. The patients who draw the diagonal of Figure 5 toward the half-circle are more often characterized as immature, passive, inadequate, or dependent, and more often exhibit hostile or aggressive behavior. It is interesting to note that the patients in the control group who exhibit hostility are more likely to be capable of direct expression. More evidence for the apparent inability of the experimental group to face directly the source of their frustrations is seen in the fact that they show a significantly greater incidence of blackouts or amnesia. The affect of this group also seems more seriously disturbed, as seen in the greater frequencies of depression, tearfulness, and suicidal thoughts or attempts. That frustrations may always have been present for these people is suggested by the finding that serious deficit or abnormality in the home situation during formative years is mentioned with significantly greater frequency in the experimental group.

Differences appear also in the fears and dreams which reach sufficient intensity as to be included in the discharge summary. These fears or dreams all seem to represent forces of a violent nature impinging upon the patient, as illustrated by the following typical statements: "people planned to shoot him," "he felt there were guns pointing at him in the dark," "dreams of sharks," "soldiers coming at him," "feeling of a gun being discharged into his abdomen," etc.

As a very crude measure of the over-all extent of involvement, a summation was made of the number of complaints and the listed personality traits. It was found that the control group was more likely to have zero, only one, only two, or three or less such tallies, while the experimental group was much more likely to have six or more or seven or more tallies.

Discussion

It appears from the above experiment that differences in personality and even in certain personal-history items are reflected in the manner in which the subject draws the component parts of a Bender-Gestalt figure. It is especially significant that this should be true even when a very specific behavior on a single B-G figure (independently of quality of the performance) is correlated with a criterion as brief and variable in its coverage of data as a discharge summary.

It should be pointed out, however, that although statistically significant differences occur, many of these variables show a large enough incidence in both groups as to reduce the practical importance of such variables. It may be more useful, therefore, to consider a syndrome or combination of attributes by which to differentiate various groups. On the basis of this experiment such a behavior pattern might be postulated for patients drawing Figure 5 as described. In general, these patients appear immature and dependent, to have somewhat less capacity for adjustment than their ward-mates. They are prone to feel overwhelmed by their problems and to be easily frustrated. Direct attack upon the sources of this frustration seems too threatening; other less adaptive mechanisms are resorted to, such as blackouts, amnesia, and the many self-punishing somatic reactions. The frustrations experienced as a result of their inadequacy may lead to hostile and aggressive expression, which is more likely to be indirect or poorly controlled. The ineffectiveness may appear insurmountable so that there occurs a greater frequency of depression, tearfulness, and even suicide. Mention has already been made of the greater percentage of "good" prognoses in the control group. Though significant only at the .10 level, it is noted that there is a greater tendency for the patients within this syndrome to be completely incapacitated or to have been judged incompetent by a court. As pointed out earlier, the larger proportion of grossly unfavorable factors in the home environment is suggestive in the consideration of etiological elements.

For those clinicians interested in the symbolic value of the B-G figures, the statement might be made that the direction in which the projection in Figure 5 is drawn may represent the impinging forces within the patient's environment, which are perceived as overwhelming and punishing, as contrasted with other patients whose attack is directed outwardly (the same direction in which they draw the projection on Figure 5) against their problems.

What defenses are utilized in the individual case cannot be determined by knowledge of this one relatively minor aspect of B-G behavior but are dependent upon other findings on the test. It is hoped that research now in progress using a configural approach with objectively scored factors may increase the usefulness of the B-G as a clinical and experimental instrument. This experiment seems to indicate that the order in which lines are drawn and the directionality of such lines should be included in B-G scoring systems.

Summary

Comparison was made between 75 randomly selected neuropsychiatric patients and 75 patients from the same population who were known to have drawn the diagonal projection on Figure 5 of the B-G toward the inverted half-circle of dots, rather than in the popular manner, i.e., outwardly from the half-circle. Overt behavior and traits of the patient and certain personal-history items as described in the discharge summary were the variables compared in the two groups. The significance of the differences of percentages of occurrence of each attribute were tested using phi coefficients. Twenty-seven variables were found to have differences significant beyond the .01 level of confidence, while nine variables were significant at the .05 level. On the basis of these findings it was suggested that the order and directionality used in drawing parts of the B-G figures be included in data derived by testing with that instrument.

Received September 22, 1952.

References

1. Bender, L. A. A visual motor Gestalt test and its clinical use. *Amer. Orthopsychiat. Ass. Res. Monogr.*, 1938, No. 3.
2. Billingslea, F. Y. The Bender-Gestalt: an objective scoring method and validating data. *J. clin. Psychol.*, 1948, **4**, 1–27.
3. Hutt, M. A. A tentative guide for the administration and interpretation of the Bender-Gestalt test. U. S. Army Adjutant General's School, 1945 (restricted).
4. Jurgensen, C. E. Table for determining phi-coefficients. *Psychometrika*, 1947, **12**, 17–29.
5. Pascal, G. R., & Suttell, Barbara J. *The Bender-Gestalt test.* New York: Grune & Stratton, 1951.
6. Peek, R. M., & Quast, W. *A scoring system for the Bender-Gestalt test.* Hastings, Minn. (Box 292) and Minneapolis, Minn. (2810 42nd St.): Authors, 1951.

Journal of Consulting Psychology
Vol. 17, No. 3, 1953

Test-Retest Stability of MMPI Scales for a Psychiatric Population[1,2]

Albert Rosen
Veterans Administration Hospital, Minneapolis, Minnesota

The study of reliability of personality tests involves some problems which are not present in the fields of intelligence and achievement testing. Reliability defined as accuracy of measurement or in terms of a low standard error of measurement cannot be evaluated adequately for personality tests. Internal consistency measures of reliability are not appropriate for scales such as those in the MMPI [1]. Some of these scales demonstrate considerable inhomogeneity, for they were not constructed with the intention of achieving internal consistency. Thus, although the test-retest method is more suitable for assessing reliability, the temporal changes in test-retest data make it more desirable to refer to "stability" rather than "reliability."

There has been some question as to whether a high degree of reliability is desirable in personality tests [3]. According to this latter argument, the sensitivity of a personality scale to changes in an individual, say, through psychotherapy, is much more important than reliability. In intelligence and achievement testing, on the other hand, the ideal test is one which gives consistent results, regardless of extraneous factors.

Perhaps the situation might be clarified if one were to think in terms of both stability and sensitivity. Stability would be measured by the consistency of test results over a period of days. Sensitivity would be evaluated by the lack of consistency of scores after a period of time, following some change which had occurred in the individual as a result of psychiatric therapy or gross environmental alteration. Both of these conditions are desirable features of a personality scale. A score for an individual on any scale at a given time should be within a reasonable distance of his "true" score. On the other hand, the scale should have the capacity for reflecting modifications in personality with the advent of radical change in environmental status.

The purposes of the present study were (*a*) to measure the stability of the individual MMPI scales for a psychiatric population over a brief test-retest interval by means of product-moment correlation coefficients, and (*b*) to evaluate what changes occur in patients' MMPI scores as a result of hospitalization and participation in the hospital regime for a few days before they have received much, if any, therapy of a formal nature.

Selection of Subjects

Every new, testable patient who was admitted to the psychiatric section of the Minneapolis VA Hospital during an eight-week period was given the MMPI. All patients are males. Retesting was done within two to seven days until forty test-retest cases were obtained. It was decided beforehand to exclude from the sample the following cases:

1. Patients retested more than seven days after the initial test. The seven-day limit was arbitrarily set in order to reduce the effects of psychotherapy on scores.

2. Former patients. Each of these patients had had one or more tests during one or more previous admissions. It was anticipated that sufficient re-

---

[1] Reviewed in the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the author are a result of his own study and do not necessarily reflect the opinion or policy of the Veterans Administration.

[2] Read at the annual meeting of the American Psychological Association, Division of Clinical and Abnormal Psychology, Washington, D. C., September, 1952.

sistance to retesting would be encountered in these patients to provide an inadequate measure of stability.

3. Soldiers. Men in active service, when admitted to the hospital, often possess strong motivation to take the test in such a way as to give a very poor psychiatric picture, with the hope of obtaining discharge from the service.

4. Patients given the group form of the MMPI as the initial test. It was decided to restrict the study to patients who had taken the individual (card) form because of any differences which might occur as the result of the patient taking two somewhat different tests.

Of the 89 patients admitted to the hospital during the eight-week period, the records of 49 cases could not be used in the study because of the four reasons described above, or because a test and retest were not available. The reasons that these 49 cases were excluded are classified below, with the number of cases indicated for each category.

1. Initial test unavailable (12): (*a*) In Psychiatry too brief a period to be tested (5), (*b*) Too disturbed to take test (5), (*c*) Received electroshock therapy almost immediately (2).

2. Retest MMPI unavailable (37): (*a*) Former patients (14), (*b*) Initial MMPI given more than one week prior to admission to Psychiatry (6), (*c*) Not retested within a week of the initial test (4), (*d*) Group form administered as initial test (3), (*e*) Soldier (2), (*f*) Electroshock therapy given before retest (2), (*g*) Refused retest (2), (*h*) Too disturbed for retest (2), (*i*) Left Psychiatry before retest (2).

Procedure for Administration of Tests and Retests

Each patient was individually administered the initial test within a few days after admission by the psychologist on the psychiatric team to which the patient was assigned. Thus, only a few cases were given the initial test by the writer. Retesting was done within two to seven days by the writer. In most cases the retest was given individually, but in approximately eight cases the patients were seen in pairs, and one group of five patients was tested at a time.

Each patient was given the same rationale for the retesting, although no attempt was made to word it exactly in the same way. Each patient was told, in effect, that the retest would be of value to the staff and the patient in affording a dual base line for comparison with any later psychological evaluations which might be desired for assessing the patient's status. However, the patient was told to answer the items as he felt at the time. The latter instruction was designed to counteract any belief on the patient's part that he might be expected to duplicate his initial test responses.

Table 1

Test-Retest Time Intervals for the 40 Patients in the Stability Study

	Number of Days from—		
	Admission to first test	First test to second test	Admission to second test
Q_1	1.3	3.1	5.5
Md	2.8	3.9	7.0
Q_3	5.2	4.8	9.8
M	3.6	4.1	7.7
Range	0–12	2–7	2–19

In Table 1 are shown the summary statistics on the number of days from (*a*) date of hospital admission to initial test, (*b*) initial test to retest, and (*c*) hospital admission to retest. The Christmas vacation period intervened during this study, so that some of the patients were seen for initial testing later than would normally occur. Four patients took the first test more than seven days after admission. On the average, however, patients were tested within three days of admission and retested within four days. No patient was retested within forty-eight hours following the initial test in order to avoid contaminating the results with a memory factor.

Description of Patients in the Stability Study

According to primary psychiatric diagnosis, the sample was comprised as shown in Table 2. Data were also recorded on each patient for age, occupational level, years of formal education, marital status, and intelligence. Medians for the group are as follows: Age—32.4 years; education—11.2 years; occupational level—4.4.[3] Of the total group, 32.5% had never been married, and 53.8% possessed an equivalent Wechsler-Bellevue IQ of 110 or higher, as determined from Shipley-Hartford test results. The stability sample demonstrated no significant difference on any of these vari-

---

[3] The occupational rating scale devised by Warner, Meeker, and Eells [4] classifies occupations in levels 1 (the highest) through 7.

Table 2
Primary Diagnoses of the Patients in the MMPI Test—Retest Sample

Diagnosis	Number in Sample	Per cent of Sample
Anxiety reaction	12	30.0
Conversion reaction	3	7.5
Depressive reaction	9	22.5
Somatization reaction	1	2.5
Neurosis total	25	62.5
Paranoid schizophrenia	5	12.5
Schizophrenia, unclassified	3	7.5
Paranoid state	2	5.0
Involutional melancholia	1	2.5
Psychosis total	11	27.5
Inadequate personality	1	2.5
Alcoholism	2	5.0
Convulsive disorder	1	2.5
Other diagnosis total	4	10.0
All cases	40	100.0

ables, nor on diagnosis, when compared with a sample of 250 patients-in-general. This latter group represents a cross section of all patients at the same hospital over a two-year period who had taken the MMPI. Therefore, results obtained in this study are most likely representative of the patient population.

Test-Retest Stability Coefficients

For each of the scales, a product-moment correlation coefficient was computed from the test and retest records of the 40 patients in the stability sample. There has been only one other study, that of Holzberg and Alessi [2], in which the stability of the MMPI scales was determined for a psychiatric group over a brief test-retest period. However, that investigation differed from the present one in the following ways:

1. One-half of their total group of 30 cases received the complete test of 550 items as the initial test, and were retested within one or two days with only the 350 items which are scored in the original scales.
2. The other 15 cases received the long and short form in reverse order within the same time interval.
3. Exactly two-thirds of their cases were psychotics as compared to only 27.5% in the present study.
4. Their sample apparently consisted of state hospital psychiatric patients of both sexes.

The product-moment stability coefficients of the two studies are shown in Table 3. The coefficients from the two studies generally correspond except on *L*, *F*, *Hs*, and *Pd*. Perhaps the coefficients for *L* and *F* are lower in the present study because of the reduced effect of the memory factor over the longer test-retest interval. Any other divergence in the results for the two studies may be due to the differences in diagnostic composition of the two samples. The MMPI scale scores of the Holzberg and Alessi sample of state hospital patients are all considerably lower than those found for the veterans in the present study. The latter are no doubt suffering from comparatively acute rather than chronic mental disorders.

Table 3
Test-Retest Stability Coefficients for MMPI Scales

Scale	Present Study ($N = 40$)	Holzberg & Alessi [2] ($N = 30$)
L	.62	.85
F	.81	.93
K	.65	
Hs	.85	.67
Hs + .5 *K*	.86	
D	.80	.80
Hy	.88	.87
Pd	.88	.52
Pd + .4 *K*	.87	
Mf	.64	.76
Pa	.75	.78
Pt	.80	.72
Pt + 1.0 *K*	.88	
Sc	.83	.89
Sc + 1.0 *K*	.86	
Ma	.56	.58
Ma + .2 *K*	.55	
Si	.83	

Note—Coefficients of .31 and .36 are significantly different from zero at the .05 level, for *N* of 40 and 30, respectively.

In the present study, the coefficients for *L* and *F* no doubt underestimate the degree of relationship between the test and retest records, because of the positive skewness in the *L* and *F* distributions of scores. On the clinical scales, the correlation coefficients range from .55 to .88. All the correlations of the clinical scales in common use are equal to or above .80, except *Pa* and *Ma* + .2*K*. The MMPI scales,

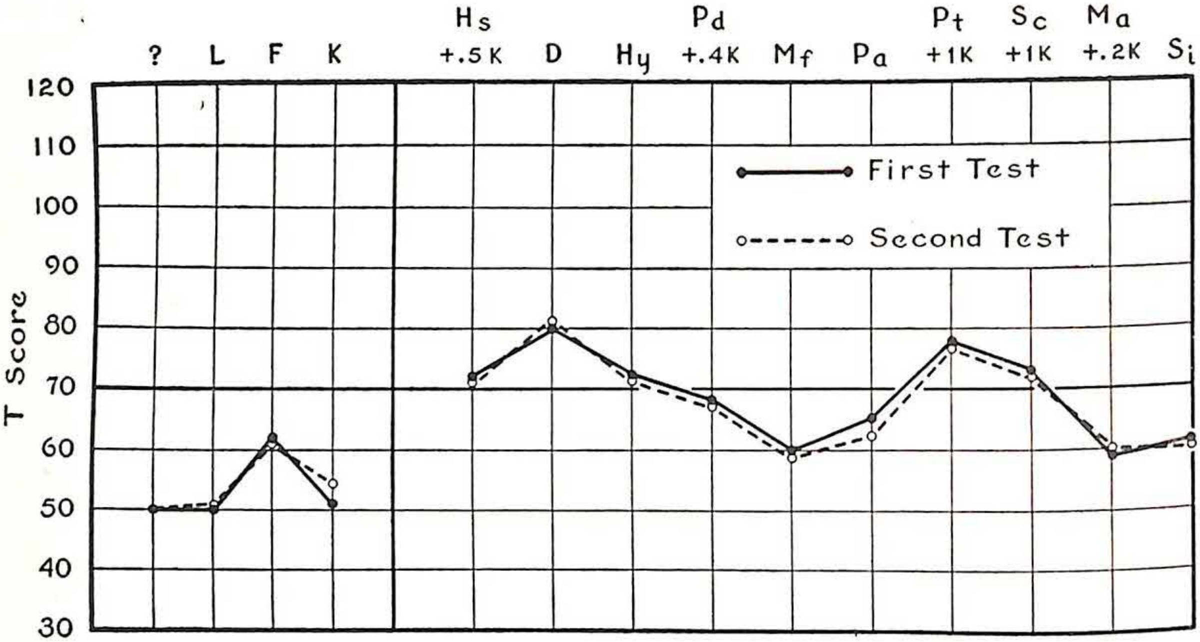


Fig. 1. Mean MMPI T-score profiles for the first and second tests ($N = 40$).

in general, demonstrate a fairly high degree of stability in an absolute sense, and a high degree of stability relative to what one would anticipate in personality scales.

It is important to consider the factors which may have influenced the degree of consistency obtained in this study. The following conditions of administration were not uniform for initial test and retest, and would tend to reduce stability.

1. Variations in motivation from test to retest. Some of the patients took the second test somewhat reluctantly because they were not strongly convinced that it would be of any additional value.
2. Individual administration of first test and group administration of retest for some cases.
3. Administration of test and retest by different examiners.

On the other hand, the second test was given within two to seven days of the first, so that memory for some of the items may have contributed in the direction of enhancement of stability. It is difficult to determine how much weight can be assigned to each of these conditions. The first is by far the most important factor. Probably the total effect was to somewhat underestimate stability.

Changes in Means from Test to Retest

Clinicians often speculate about what changes occur in patients merely as a result of hospitalization and participation in the hospital regime for a few days before they have received much, if any, formal therapy. The data from this sample provide this information, as it is measured by the MMPI. Table 4 presents the

Table 4

Differences in Raw Score Means on MMPI Scales from Test to Retest ($N = 40$)

Scale	Mean First Test	Mean Retest	Difference	Standard Deviation First Test	Standard Deviation Retest
L	4.05	4.35	.30	2.17	2.09
F	7.78	7.48	–.30	5.12	4.75
K	13.08	14.58	1.50*	4.08	4.69
Hs	13.00	11.90	–1.10	6.25	7.26
Hs + .5 *K*	19.82	19.45	–.37	6.25	6.90
D	29.20	29.55	.35	6.27	6.66
Hy	28.50	27.62	–.88	6.38	7.71
Pd	21.50	20.45	–1.05*	5.49	5.50
Pd + .4 *K*	26.65	26.28	–.37	4.86	4.97
Mf	25.48	24.75	–.73	4.20	4.21
Pa	13.00	11.92	–1.08*	4.39	4.04
Pt	23.12	21.02	–2.10*	9.45	9.57
Pt + 1.0 *K*	36.20	35.50	–.70	7.98	8.05
Sc	21.40	18.95	–2.45*	10.94	9.90
Sc + 1.0 *K*	34.48	33.43	–1.05	9.49	8.49
Ma	17.88	18.35	.47	4.30	5.28
Ma + .2 *K*	20.52	21.25	.73	3.87	4.90
Si	34.48	33.25	–1.23	11.00	11.08

*Difference significant at the .05 level.

differences in raw score means for the 40 patients in the stability study who were, on the average, tested within three days of admission to the hospital and retested four days after the first test. Figure 1 portrays the test and retest mean T-score profiles based on the raw scores in Table 4.

There are five significant differences in means from the first test to the retest. The increase in *K* score and the decrease in *Pd, Pa, Pt,* and *Sc* point to a movement in the "normal" direction with an improvement in defensive structure. After a few days of hospitalization, there is a general over-all decrease in MMPI scores for the group as a whole.

Summary

1. A group of 40 male, veteran, psychiatric hospital patients were administered the individual form of the MMPI on an average of three days after admission to the hospital and again on an average of four days after the initial test.

2. For clinical scales in common use, test-retest stability coefficients are almost all between .80 and .88.

3. There is an over-all tendency for reduction in MMPI scale scores and increase in defensiveness during the first few days of hospitalization before much, if any, formal therapy has been initiated.

Received September 22, 1952.

References

1. Hathaway, S. R., & McKinley, J. C. *Manual for the Minnesota Multiphasic Personality Inventory.* (Rev. Ed.) New York: Psychological Corp., 1951.
2. Holzberg, J. D., & Alessi, S. Reliability of the shortened MMPI. *J. consult. Psychol.,* 1949, **13,** 288-292.
3. Horn, D. Intra-individual variability in the study of personality. *J. clin. Psychol.,* 1950, **6,** 43-47.
4. Warner, W. L., Meeker, M., & Eells, K. *Social class in America: A manual of procedure for the measurement of social status.* Chicago: Social Science Research Associates, 1949.

Journal of Consulting Psychology
Vol. 17, No. 3, 1953

Perceptual Learning and Age

Bernard Hanes[1]
Ohio State Penitentiary

Gilbert [1] demonstrated that relationships between learning and age are probably a function of the method of presentation and measurement. After matching old and young individuals on a vocabulary test he concluded that the least amount of deficit occurs in the immediate recall of brief new material; the greatest deficit occurs in the recall, after brief exposure, of more complex material.

Gilbert's study also demonstrated that memory involving the formation of new associations was extremely difficult for senescents, thus pointing to a relative intellectual inflexibility for this age group.

Halstead [2] approximated the same conclusions in his study of seniles aged 69 through 83. These old people were best at rote memory, immediate visual recognition, and doing tasks involving early acquired habit patterns. Remembering or learning involving the reversal of old habits was considered to be poor.

Ruch's [3] work assumes that later maturity should bring a differential set of changes in rate of learning. He discovered that older individuals were relatively more handicapped if the new learning required the attachment of an antagonistic response to a previously learned S-R connection than if the response was consonant with past experience. Relatively meaningless responses stood intermediate in difficulty of learning between the consonant and the dissonant material. His measure of the effects of the type of material was in terms of ability to form S-R connections. He used, of course, a critical score as his measure of performance.

It seems entirely possible that this learning deficiency is in part due to errors in perception of the material when it is presented. This deficiency might be more marked for the dissonant than for the consonant material and, in turn, more marked for the older than for the younger groups.

Exposure Materials Used

In order to obtain a clearer view of this problem, a task consisting of tachistoscopically presented material was chosen for its relationship to the past experience of the subjects. The task consisted of three types of material.

Type A: This is considered as highly stable, unambiguous material, essentially agreeing with past experience, "true material," as follows: $10 \times 7 = 70$, $23 - 7 = 16$, $1 \times 12 = 12$, $16 \times 3 = 48$, $6 \times 12 = 72$, $13 - 2 = 11$, $12 \times 5 = 60$, $11 \times 9 = 99$, $15 - 5 = 10$, and $21 - 2 = 19$.

Type B: This material is considered as nonstable, requiring a reversal of past habit associations, and is essentially contrary to past experience, "false material," as follows: $20 \times 4 = 10$, $9 \times 12 = 90$, $18 - 9 = 11$, $9 \times 14 = 36$, $20 - 5 = 19$, $33 - 7 = 24$, $12 \times 0 = 12$, $4 \times 12 = 49$, $32 - 32 = 1$, and $55 - 5 = 49$.

Type C: This material is considered lowgrade, rather nonsensical information, requiring new associations, and is essentially nonsense, as follows: $14 : 3 = 12$, $21 : 17 = 9$, $5 : 15 = 62$, $19 : 49 = 7$, $22 : 18 = 7$, $8 : 31 = 53$, $23 : 5 = 17$, $32 : 19 = 6$, $54 : 1 = 85$, and $29 : 6 = 53$.

Hypotheses Tested

The hypotheses tested were: (*a*) Highly stable, unambiguous information should be relatively intact in later maturity. (*b*) Nonsense material which calls for new associations should show an intermediate decline in old age. (*c*) Information contrary to past experience should show the greatest decline in later maturity.

Description of the Subjects

One hundred and eighty white, recent arrivals at the Ohio State Penitentiary were each individually told that they composed a part of a special group of high-caliber men, and that we wished to compare them on a test with other inmates. These subjects were not psychotic and had normal scores on the Minnesota Multiphasic Personality Inventory. Their IQ scores on the revised Beta examination ranged from 91 to 123. The subjects all agreed they would do their best on the test; in fact, they were quite anxious to get started.

---

[1] Now in Dallas, Texas.

Although it is possible to state theoretically the criteria of functional old age [4], exact means of measuring these criteria have not, as yet, been developed. Therefore, in this study, it was deemed appropriate to use chronological age as the criterion of the aging process. Accordingly, the following terms were used in this study: (I) "Young Group" was used to designate the age period from 20 to 34 years of age. (II) "Middle Maturity Group" was used to designate the age period from 35 through 49 years of age, and (III) "Later Maturity" or "Old Group" was used to designate the age period from 50 to 70 years of age. Each of these groups was composed of 60 individuals of approximately average intelligence and of normal visual acuity.

Experimental Design

Tachistoscopic exposures of thirty 3″ × 5″ white index cards containing the various types of information discussed above were randomly presented. The exposure time was 1/10 second. The number of correct perceptions constituted the score.

Sixty subjects were assigned to each of the various groups as described above. Group I was utilized as a control group, while Groups II and III were utilized as experimental groups. The subjects were individually called into the testing room, and seated in front of the Renshow tachistoscope. The practice trials consisted of tachistoscopic exposure of one of each of the various kinds of material. The following instructions were given to each subject.

> Here is a task in which you look into the instrument you see in front of you, like this [demonstrate]. Here is some idea of what to expect. You may see something like this, [show one of the practice cards], or you may see something like this [show another practice card]. If you aren't quite sure of what you see, go ahead and guess. Before running through the regular test cards, I will let you have three practice trials. [After each practice card had been "flashed" the subject was shown exactly what information was contained on the card.] You have now finished the practice test. Do your best, and tell me exactly what you see.

Statistical Analysis of Results

The results were submitted to an analysis of variance. All the *F* ratios in Table 1 are significant. However, in order to determine where these significant differences lie, the means, standard deviations, and *t* ratios of the various data were computed, and are shown in Table 2. The means of the various groups for each kind of material are shown graphically in Figure 1.

Table 1
Analysis of Variances

Source of variation	*df*	Mean square	*F*
Group I			
Between types of material	2	389.73	125.31
Between people	59	7.92	2.55
Residual (error)	118	3.17	
Group II			
Between types of material	2	441.76	158.34
Between people	59	6.96	2.49
Residual (error)	118	2.79	
Group III			
Between types of material	2	408.10	124.04
Between people	59	8.03	2.44
Residual (error)	118	3.29	

Table 2
Means, *SD* and *t* Ratios for the Various Materials

Group number	M	*SD*	Differences	*t* ratios
III				
A	7.00	2.08	A-B	10.74
B	2.87	2.19	A-C	13.90
C	2.34	2.19	B-C	3.76
II				
A	8.17	1.76	A-B	10.40
B	4.40	2.00	A-C	11.71
C	2.90	2.28	B-C	1.28*
I				
A	8.52	1.64	A-B	27.94
B	4.58	2.34	A-C	32.08
C	3.58	2.49	B-C	2.23

* Nonsignificant at 5% level of confidence.

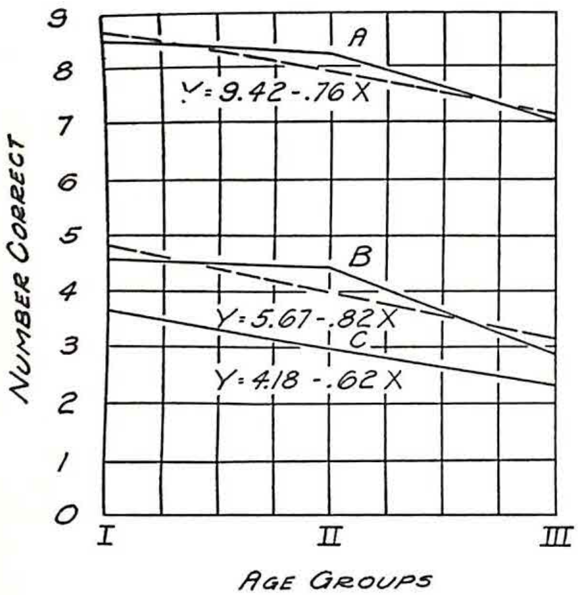


Fig. 1. Comparison of the three types of materials in each age group.

Discussion

The analysis of variances reveals a real difference between types of material exposed and between individuals within each age group.

An interaction is probably demonstrated by the nonsignificant *t* ratio of 1.32 on material *C*, between Groups III and II. Perhaps this interaction pertains to a relaxation process taking effect in the ages after 35.

Judging from Figure 1, a deficit following a definite straight line from young to old occurs in material *C*. The dotted lines, constructed by the least-square method, show the deficit that should occur on materials *B* and *C*. The line for material *C* is, of course, identical with the least-square line of best fit.

The amount of retardation in deficit in materials *A* and *B* is theoretically demonstrated by the vertical space between the dotted and solid lines. Evidently, the largest deviation of materials *A* and *B* from the general trend occurs in the middle age group and grows impreceptibly less as the extreme ages are approached.

No particular relationship is demonstrated between kinds of material and age. The slopes of the various regression lines are practically identical and slight variations from parallel are treated as error. Thus, the study shows a relationship between performance and age, with all three groups experiencing greater difficulty in learning the nonsense material. Apparently then, materials requiring the reversal of old habits show a substantially greater deficit in the aged, but the greatest deficit is manifested in materials calling for the learning of new associations. Hypotheses (*b*) and (*c*) are untenable.

Table 3

t Scores between Groups on Various Conditions

Groups	*t* ratios under Condition A	B	C
III–II	3.26	3.92	1.32*
III–I	4.27	4.16	2.82
II –I	1.11*	.44*	1.56*

* Nonsignificant at 5% level of confidence.

Conclusions

Decline appears to be most apparent in learning requiring new associations, rather than learning which demands a reorganization of previously formed habit patterns. No evidence for any relationship between kinds of material and decline was found; but, rather a consistent relationship between age and performance was discovered. The younger person performs better, but experiences approximately the same difficulty with the various types of materials as the older person.

Received October 8, 1952

References

1. Gilbert, Jeanne G. Mental efficiency in senescence. *Arch. Psychol., N.Y.*, 1935, No. 188.
2. Halstead, H. A psychometric study of senility. *J. ment. Sci.*, 1943, **89**, 363–373.
3. Ruch, F. The differential effects of age upon human learning. *J. gen. Psychol.*, 1934, **11**, 261–286.
4. Stieglitz, E. J. *Geriatric medicine.* Philadelphia: Saunders, 1943.

Journal of Consulting Psychology
Vol. 17, No. 3, 1953

Wechsler-Bellevue Split-Half Subtest Reliabilities: Differences in Age and Mental Status[1]

Jack Botwinick
Section on Gerontology, National Heart Institute
National Institutes of Health, PHS, FSA, Bethesda, Maryland
and the Baltimore City Hospitals
Baltimore, Maryland

Considering the extent to which the Wechsler-Bellevue test [13] is used and the various groups to which the test is administered, the number of studies investigating the reliability of the test is limited [10]. Derner, Aborn, and Canter [2], and Webb and De Haan [12] have examined "normal" subjects, while others [4, 5, 8, 9] have utilized psychiatric subjects. Webb and De Haan, who investigated subtest reliability in a paranoid schizophrenic group as well as in a "normal" one, and Gilhooly [4], who investigated the reliability of only 4 verbal subtests, computed coefficients of reliability by the split-half method. The other investigators computed reliability by the retest method.

Rabin and Guertin [10] reported that the studies utilizing psychiatric subjects are too inconclusive because the retest reliabilities were confounded with clinical changes and with practice effects. It was for this reason, plus the fact that the Wechsler-Bellevue test is usually administered in one test period, that Webb and De Haan believed it important to compute reliability by the split-half method.

Webb and De Haan [12] reported higher split-half reliability coefficients in the performance subtests for their paranoid schizophrenic group than for their "normal" group, although only one difference between the coefficients was significant at the 1 per cent level and one at the 5 per cent level. Helmick [6] pointed out that since the magnitudes of reliability coefficients are dependent upon the variability in the groups, the conclusions reached by Webb and De Haan were unwarranted. He recalculated the data of Webb and De Haan and found that when the respective variabilities were allowed for, the magnitudes of the reliability coefficients were about the same for the paranoid schizophrenic and the "normal" groups.

The present study was designed to extend our knowledge of Wechsler-Bellevue split-half reliabilities.

Procedure

Comparison was made between the subtest reliabilities computed by Webb and De Haan for a "normal" group [12] with the subtest reliabilities of a "normal" but older group. In addition, comparison was made between subtest reliabilities of the older "normal" group and those of a matched group comprised of hospitalized patients diagnosed as "senile psychosis" and "psychosis with cerebral arteriosclerosis."

Subjects of the older "normal" group were originally examined for the purpose of another study by Fox and Birren [3]. The mean age of this sample population was 64.3 years, $\sigma = 3.19$, and the mean age of the sample population of Webb and De Haan [12] was 37.6 years, $\sigma = 11.20$. The former group will henceforth be referred to as the "old control" and the latter as the "young control." The mean IQ of the old control group was 100.8, $\sigma = 11.4$, and of the young control, mean IQ $= 97.8$, $\sigma = 11.07$. Both groups consisted of 50 white subjects who were apparently free of aberrant mental conditions. The only determinable difference between these two groups, in addition to the above, was a sex difference. The young control group consisted of women whereas the old control included both men and women. The hospitalized group that was compared with the old control group was originally examined for the purpose of a validity study [1]. It was indicated that the old control group and the hospitalized group were comparable except for mental status. The hospitalized group will be referred to as the "senile group." It included 31 male and female subjects.

---

[1] The criticism and assistance of N. W. Shock in the preparation of this report is gratefully acknowledged. Mrs. E. Benser gave valuable aid in the computational work.

The method of Webb and De Haan was used for estimating the reliability coefficients. The scores on odd versus even items of the subtests were correlated and reliability was estimated by the Spearman-Brown formula for double length of test. The reliability of the Digit Span subtest was determined by correlating Digits Forward and Digits Backward instead of correlating odd versus even items. The reliability coefficient of the Object Assembly subtest was estimated by intercorrelating the three items, transforming the intercorrelations to z values, and retransforming the mean z value to the correlation coefficient. The reliability coefficient was estimated by the Spearman-Brown formula for triple length. The reliability coefficient for the Digit Symbol subtest was not determined since the total subtest is measured by time and data were not available for individually timed items.

Mean differences in coefficients of reliability between old and young control groups and between old control and senile groups were determined by z-value transformations.

To allow for the differences in variability between the groups, the method suggested by Helmick [6] was used. That is, the standard error of measurement [7, p. 130] was computed for each subtest in the groups and F tests were applied to the standard errors of measurments between groups. In addition, the coefficients of reliability were corrected for differences in range [7, p. 134] and the corrected reliabilities were compared by z-value transformations.

Results

The corrected subtest reliability coefficients of the old control and senile groups are presented in Table 1 (columns marked r). One reliability, that of the Comprehension subtest of the old control group ($r = 0.35$) is so low as not to be significantly different from zero ($p < .01$). Three coefficients prior to correction by the Spearman-Brown prophecy formula were also not statistically greater than zero ($p < .01$): the Comprehension and Object Assembly subtests of the old control group, and the Digit Span subtest of the senile group. Webb and De Haan [12] reported that the lowest four uncorrected coefficients for the young control sample population were not significantly above zero at $p < .01$, and that the corrected coefficient for the Picture Arrangement subtest was below this level.

Table 1

Age Differences in Reliability and Standard Error of Measurement of Subtests of the Wechsler-Bellevue Scale

Subtests	A Young control* $N = 50$		B Old control $N = 50$		C Senile $N = 31$	
	r	σ meas.	r	σ meas.	r	σ meas.
Vocabulary†	.94	1.37	.96	1.48	.94	1.69
Arithmetic	.82	1.06	.70	1.11	.92	0.85
Information	.82	1.58	.90	1.21	.93	1.59
Block Design	.76	2.90	.84	2.51	.89	2.14
Similarities	.74	1.80	.89	1.47	.83	1.43
Comprehension	.53	1.96	.35	2.22	.69	1.89
Object Assembly‡	.46	1.90	.62	2.50	.66	2.82
Digit Span§	.44	1.45	.76	1.07	.51	1.44
Picture Completion	.42	1.48	.83	1.25	.82	1.15
Picture Arrangement	.29	2.23	.55	1.57	.87	1.18

* From data of Webb and De Haan [12].

† N = 36 for Column A and N = 34 for Column B of the Vocabulary subtest.

‡ N = 37 for Column A of the Object Assembly subtest. Object Assembly reliability coefficients were estimated by the average intercorrelations of the items.

§ N = 49 for Column A of the Digit Span subtest.

When subtest reliability coefficients were compared between control groups, i.e., between young and old control, only the difference in reliability of the Picture Completion subtest was significant at the 1 per cent confidence level. The remaining subtest reliabilities were not significantly different at the .01 level although the reliability coefficients of the Digit Span subtest and the Similarities subtest were significantly different at $p < .05$. Significant differences in reliability coefficients between old control and senile groups were seen in the Arithmetic and Picture Arrangement subtests ($p < .01$). The difference in reliability of the Comprehension subtest was significant at $p < .05$ but no significant differences in reliabilities were indicated even at this confidence level for the other subtests.

The differences between the old control and senile groups with respect to the standard errors of measurements presented in columns marked σ_{meas} in Table 1 were not statistically significant. Age differences in the standard errors of measurements were significant for the Picture Arrangement subtest at the 1 per cent level and for the Digit Span subtest at the 5 per cent level.[2]

When the reliability coefficients were corrected for range and compared by z-value transformations, a difference between the two elderly groups was found at the 5 per cent level for only the Information subtest and no statistically significant differences, even at this confidence level, were found for the remaining subtests. When age comparisons were made for the reliability coefficients corrected for range, only the Object Assembly and Picture Completion subtests were found to be significant. The confidence levels were 5 and 1 per cent respectively when the range of the old control was corrected to that of the young control, and 1 and 5 per cent respectively when the range of the young control was corrected to that of the old control.

Discussion

The results of this investigation have indicated that the range of the coefficients of reliability for the different subtests was large in all groups. From Webb and De Haan's data [12], differences of .65 and .37 between the highest and lowest reliability coefficients were computed for a "normal," i.e., young control, and a paranoid schizophrenic group respectively. The present data indicated ranges of .61 and .43 for the old control and senile psychotic groups respectively. In general, the larger reliability coefficients were about of the same magnitude for all groups, but the lower coefficients appeared to increase in magnitude with age and psychoses. This was also reflected in the direct comparisons of reliability coefficients between the two control groups and between the control and psychotic groups of similar age. Webb and De Haan reported that the subtests in which the paranoid schizophrenic group were inferior were the ones in which higher reliabilities were manifested. In the present data, performance was lower in the senile group than in the old control group for all subtests. Yet those subtest reliabilities that were significantly different were in all instances higher in the psychotic group than in the old control group.

Because of the relationship between reliability coefficients and range of ability, these findings suggested that there were group differences with respect to variability, and that variability increased with age and psychoses. Helmick [6] reported that in all but one subtest, Webb and De Haan's paranoid schizophrenic group were more variable than the young control group, although only one difference in variance was significant at less than the 5 per cent level. In the present data, the old control group was more variable than the young control group in seven of the ten subtests, although only the Object Assembly and Picture Completion subtests were significantly different with respect to variance ($p < .01$). The senile group was more variable than the old control group in six of the ten subtests, but three subtests were significantly different with respect to variance (Arithmetic and Information at $p < .01$, and Picture Arrangement at $p < .05$).

In view of these differences in variability, the standard error of measurement appeared to be a more adequate index for comparison purposes. Helmick [6] found no significant differences in the standard errors of measurements between the young control and young psychotic groups and the present data indicated no significant differences between the old control and senile groups. Therefore, the indications are that in these studies there were approximately equal errors in measurement with the Bellevue subtests for the different mental conditions within an age group. Between age groups, however, i.e., between young and old control groups, the Digit Span and Picture

---

[2] The author wishes to thank Drs. Webb and De Haan, who on request, forwarded their data for the Object Assembly subtest so that the present analysis would be complete.

Arrangement subtests were not of equal errors in measurement. There was greater error of measurement in the young control group.

Conclusions drawn from the comparisons of reliability coefficients corrected for range are tenuous. As Webb [11] indicated, this correction creates a theoretical population and has more statistical than clinical meaning. Accordingly, the differences between groups in reliability coefficients corrected for range were not given the consideration that the differences in uncorrected coefficients and standard errors of measurement were given.

The present results were congruous with those of other reliability studies [2, 4, 5, 8, 9, 12] on the Wechsler-Bellevue test. For pattern analysis and diagnostic considerations, where each subtest score is evaluated in terms of its deviation from the other scores, the error of measurement and the reliability of subtests must be improved before confidence in evaluation can be expected.

Summary

1. Three groups that were examined for the purposes of separate studies [1, 3, 12] were used. Two groups were apparently free of aberrant mental conditions and were apparently matched except for age and sex. The third group was comprised of 31 men and women hospitalized for senile psychosis and psychosis with cerebral arteriosclerosis. This group was matched with the older "normal" group and the only apparent difference was mental status.

2. The general method of computing the reliability coefficients was odd-even item correlations corrected by the Spearman-Brown prophecy formula. The split-half reliability coefficients for each subtest were compared between "normal" groups of different age and between the groups of the same age but different mental status.

3. Age comparisons in reliability coefficients indicated that a difference at $p < .01$ was found only for the Picture Completion subtest. Comparisons between groups of different mental conditions of the same age indicated differences at $p < .01$ in reliability coefficients for the Arithmetic and Picture Arrangement subtests. When significant differences occurred, poorer performance scores were found for those subtests of higher reliability. Accordingly, standard errors of measurements were computed for each subtest and comparisons between groups were made. No differences were found between the old "normal" and senile psychotic groups but the Picture Arrangement subtest was significantly different at $p < .01$ between old and young "normal" groups.

4. The coefficients of reliability and the standard errors of measurements, in general, were found to be too indicative of error to warrant confidence in making diagnostic evaluations by pattern analysis from scores on individual subtests.

Received October 13, 1952.

References

1. Botwinick, J., & Birren, J. E. The measurement of intellectual decline in the senile psychoses. *J. consult. Psychol.*, 1951, **15**, 145–150.
2. Derner, G. F., Aborn, M., & Canter, A. H. The reliability of the Wechsler-Bellevue subtests and scales. *J. consult. Psychol.*, 1950, **14**, 172–179.
3. Fox, Charlotte, & Birren, J. E. Intellectual deterioration in the aged: agreement between the Wechsler-Bellevue and the Babcock-Levy. *J. consult. Psychol.*, 1950, **14**, 305–310.
4. Gilhooly, F. M. Wechsler-Bellevue reliability and the validity of certain diagnostic signs of the neuroses. *J. consult. Psychol.*, 1950, **14**, 82–87.
5. Hamister, R. C. Test-retest reliability of the Wechsler-Bellevue. *J. consult. Psychol.*, 1949, **13**, 39–44.
6. Helmick, J. S. Reliability or variability? *J. consult. Psychol.*, 1952, **16**, 154–155.
7. McNemar, Q. *Psychological statistics.* New York: Wiley, 1949.
8. Rabin, A. I. Fluctuations in the mental level of schizophrenic patients. *Psychiat. Quart.*, 1944, **18**, 78–91.
9. Rabin, A. I. Test constancy and variation in the mentally ill. *J. gen. Psychol.*, 1944, **31**, 231–239.
10. Rabin, A. I., & Guertin, W. H. Research with the Wechsler-Bellevue test: 1945–1950. *Psychol. Bull.*, 1951, **48**, 211–248.
11. Webb, W. B. Corrections for variability: a reply. *J. consult. Psychol.*, 1952, **16**, 156.
12. Webb, W. B., & De Haan, H. Wechsler-Bellevue split-half reliabilities in normals and schizophrenics. *J. consult. Psychol.*, 1951, **15**, 68–71.
13. Wechsler, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.

Journal of Consulting Psychology
Vol. 17, No. 3, 1953

Changes in Wechsler-Bellevue Test Performance following Prefrontal Lobotomy[1]

Earl D. Markwell, Jr., William M. Wheeler
Veterans Administration Center, Los Angeles, Calif.

and Helen Kitzinger
University of Southern California

Studies of the results of using the Wechsler-Bellevue Intelligence Scale [9] on various categories of neuropsychiatric patients have been numerous since its publication in 1939. Despite the wide interest in and use of this test, few such studies have concerned themselves with the changes in the subtest scores resulting from the prefrontal lobotomy operation. Of the various articles which have reviewed research with the Wechsler-Bellevue test [6, 7, 8], only the most recent survey article by Rabin and Guertin indicates such a study. In this study McCullough [3] tested 10 patients one week before and two months after the operation. Statistically significant gains were found on the digit span and picture arrangement subtests only. Rabin and Guertin, in their review, comment on this study that the demonstrated gains attributed to the intervening therapy might have been due to practice effects. A recent article concerning test-retest reliability [1] shows that these are the two subtests with the lowest reliability, and strengthens the suggestion that the obtained results may well be due to practice.

The present study was designed to determine whether components of the IQ shift following lobotomy, even in those cases where the IQ remains relatively constant. It was hoped that such a finding would throw some light on the psychological effects of prefrontal lobotomy and consequently on the functions of the prefrontal lobes. In terms of the Wechsler, it was hoped that some cues as to the "functions" measured by each of the subtests might be gained.

Thus the study was expected to have a dual function—first, to clarify frontal-lobe function, and second, to clarify the meaning of shifts in subtest scores on the Wechsler-Bellevue. The criticism can be made that the study is circular and in a very real sense it is. The purpose of discovering concomitant variation seems quite defensible, however.

Method

Wechsler-Bellevue, Form I, weighted subtest scores obtained before lobotomy and at least four months after lobotomy were compared. In terms of standard statistical techniques, the question to be answered was: is the variation in scores obtained before lobotomy different from the variation in scores obtained after lobotomy.

A group of 17 patients was tested and retested, the retest being given between 4 and 22 months postoperatively. Fourteen of these patients were males who were between 23 and 38 years of age at the time of the prelobotomy test session. All 14 were classified as having a schizophrenic reaction; 8 as paranoid, 1 as probably paranoid, 2 as hebephrenic, 1 as catatonic, 1 as mixed, and 1, while diagnosed as a paranoid schizophrenic at the time of lobotomy, had strong psychopathic elements as indicated by both previous and later psychiatric opinion.

The three female patients ranged in age from 25 to 28 years at the time of preoperative testing, and were all diagnosed as schizophrenics: 1 as paranoid, 1 as probably paranoid, and 1 as an unclassified schizophrenic.

All of the scoring of the various tests was checked by the authors. Where the original scoring was questionable, recourse was had to the supplementary guide for administering and scoring the Form I scale [2].

---

[1] From the Veterans Administration Center, Los Angeles, Calif.

Psychiatric diagnosis and all pertinent case history data were checked for accuracy with the particular entries in the patient's general clinical case file.

Results

The means and standard deviations for the pre- and postlobotomy subtest scores for the group of 17 patients were computed. There were no statistically significant differences in scores from the two testing situations. The degree of similarity between pre- and postlobotomy subtest scores was obtained by computing correlation coefficients for each of these conditions for each of the subtests and for the verbal, performance, and total weighted scores. The limits of these correlations were then computed according to the *r*-transformation technique [4]. These results are summarized in Table 1.

Table 1

Means, Standard Deviations, and Correlation Coefficients for the Pre- and Postlobotomy Conditions

(Scores in equivalent weighted score units; *N*=17)

Test Scales	Means Pre-	Means Post-	Standard Deviations Pre-	Standard Deviations Post-	Correlation Coefficient	Correlation Coefficient Limits Max.	Correlation Coefficient Limits Min.
I	8.29	9.12	9.2	9.8	.85	.96	.53
C	6.70	7.53	7.8	8.8	.74	.92	.27
D	5.94	6.29	6.9	7.3	.17	.68	−.46
A	6.52	6.58	8.0	8.0	.89	.97	.64
S	6.76	7.82	8.2	9.0	.79	.94	.38
V	8.05	8.11	8.9	8.8	.96	.99	.86
PA	6.58	8.70	7.9	9.5	.82	.95	.45
PC	7.41	7.82	8.6	8.6	.82	.95	.45
BD	7.70	8.76	8.8	9.9	.85	.96	.53
OA	7.94	9.23	9.2	10.2	.94	.98	.79
DS	5.58	6.88	6.6	7.6	.71	.92	.22
Totals							
V	34.64	38.29	38.3	41.9	.84	.96	.50
P	35.52	41.35	40.2	44.7	.90	.97	.66
T	70.18	79.65	81.6	86.1	.71	.92	.22

Even though there were no significant differences between pre- and postlobotomy subtest scores, it was believed that a comparison of the correlation coefficient for each of the subtests where the "lobotomy variable" had intervened with test-retest reliability coefficients, where no lobotomy had intervened, would throw some light on the effects of the operation.

Two of the subtests, the digit span and the block design, were further analyzed to determine intra-subtest differences.

The means for the pre- and postlobotomy digit span subtest scores for "forward" and "backward" were computed. No differences which were statistically significant were discovered. For the block design subtest the results for Designs IV and VI, commonly believed clinically to reveal organic involvement, were separated to discover any differences between the pre- and postlobotomy conditions. Counter to expectation, no increase in the extent of organic involvement occurred following the lobotomy process. The only evidence obtained was that there was a slight but nonsignificant gain in the ability to do the easier of these block designs (IV).

The deterioration coefficients for the pre- and postlobotomy conditions were computed by using the respective subtest means for each

Table 2

Comparison Table of Pre- and Postlobotomy Correlation Coefficients and Reliability Test-retest Coefficients

Test Scales	Present Study* (*N*=17)	Rabin Study* (*N*=30)	Hamister Study* (*N*=34)
I	.85	.89	.94
C	.74	.12	.78
D	.17	.62	.63
A	.89	.75	.87
S	.79	.38	.84
V	.96	......	.90
PA	.82	.54	.78
PC	.82	.32	.68
BD	.85	.71	.67
OA	.94	.31	.62
DS	.71	.34	.79
Totals			
V	.84	.78	.91
P	.90	.52	.80
T	.71	.55	.84

Note.—Rabin and Hamister studies from Derner, Aborn, and Canter [1].

*Test-retest intervals: Present study, 4-22 mos.; Rabin study, 1-35 mos.; Hamister study, ¼-1 mo.

of these groups of scores. It was found that for the prelobotomy mean subtest scores there was an 18.77 per cent loss as compared to a postlobotomy loss of 16.83 per cent. This difference was again nonsignificant. Stated another way, the results show that there was a small, nonsignificant reduction in the value of the deterioration coefficient following lobotomy for the 17 patients tested.

Table 2 summarizes the coefficients for the two test-retest samples and for our data. It is evident immediately that the three samples are quite similar.

The most striking dissimilarities occur in the cases of the digit span subtest and the object assembly subtest. No explanation of the latter seems feasible unless we interpret it to mean that the reported reliabilities are too low for our population.

In the case of digit span, however, it is possible to hypothesize that the lobotomy operation in some way disturbs this type of function. Remembering that no significant difference between pre- and postlobotomy means was discovered (5.94 and 6.29 respectively), it is *not* possible to conclude that there is a systematic increase or decrease in some psychological function measured by the digit span subtest, but rather that this area of functioning is more affected than others by the operation in some nonsystematic fashion.

The comprehension and arithmetic subtests for the two conditions were then examined to see if any change had occurred as a result of the lobotomy in the number of items answered for each of these subtests, regardless of the quality or score received on the answer to the particular item. No statistically significant changes occurred in the patient's ability to answer the various items composing the subtest.

In an article published in 1941, Rabin [5] reported an index which he found aided the separation of schizophrenic patients from nonpsychotic patients. This index showed no change in the extent of the patients' schizophrenic reaction pattern following lobotomy.

The 17 patients used in all of the statistical computations above were drawn from a finite population of 55 patients who have had the lobotomy operation. One of the most striking effects of the lobotomy was the increase in the cases who could be tested. For the total of 55 patients, the empirical evidence indicates that the lobotomy operation did increase the number of testable patients by 20 per cent, although for 5.4 per cent of them only a partial improvement in "testability" occurred, i.e., they gave answers to some but not all of the subtests.

Summary

Pre- and postlobotomy Wechsler scores of 17 psychotic patients were compared. While a general tendency to improve was noticed, there were few significant changes. Changes in two subtests (digit span and object assembly) showed a dissimilarity with test-retest changes, as indicated by correlation coefficients. Digit span for the lobotomy group was more variable, and object assembly less variable than in the "control" condition. Thus the psychological function measured by digit span appears to be the most disturbed by the operation. Several comparisons were made (e.g., deterioration coefficient, schizophrenic index), but the most striking change was in the increase of testable patients by 20 per cent after lobotomy.

Received October 3, 1952.

References

1. Derner, G. F., Aborn, M., & Canter, A. H. The reliability of the Wechsler-Bellevue subtests and scales. *J. consult. Psychol.*, 1950, **14**, 172–179.
2. Kitzinger, Helen, & Blumberg, E. Supplementary guide for administering and scoring the Wechsler-Bellevue Intelligence Scale, Form I. *Psychol. Monog.*, 1951, **65**, No. 2 (Whole No. 319).
3. McCullough, M. W. Wechsler-Bellevue change following prefrontal lobotomy. *J. clin. Psychol.*, 1950, **16**, 270–273.
4. McNemar, Q. *Psychological statistics.* New York: Wiley, 1949.
5. Rabin, A. I. Test score patterns in schizophrenia and nonpsychotic states. *J. Psychol.*, 1941, **12**, 91–100.
6. Rabin, A. I. The use of the Wechsler-Bellevue scales with normal and abnormal persons. *Psychol. Bull.*, 1945, **42**, 410–422.
7. Rabin, A. I., & Guertin, W. H. Research with the Wechsler-Bellevue test: 1945–1950. *Psychol. Bull.*, 1951, **48**, 211–243.
8. Watson, R. I. The use of the Wechsler-Bellevue scales: a supplement. *Psychol. Bull.*, 1946, **43**, 61–68.
9. Wechsler, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.

NEW BOOKS AND TESTS

Books

Abramson, Harold A. *Problems of consciousness.* New York: Josiah Macy, Jr. Foundation, 1952. Pp. 10 + 146. $3.25.

This small book of 156 pages embodies the transactions of the third of the conferences on problems of consciousness held by the Josiah Macy, Jr. Foundation. The report is a stenographic transcript of prepared papers and the comments and discussion of the 15 participants. The interchange is not only spontaneous and interesting, but frequently revealing, and is a beautiful example of the value of verbatim recording. Editing was presumably minimal. The papers were on consciousness and brain metabolism by Seymour Kety, on hypnotic phenomena by Lewis Wolberg, and on experimental work on sleep and other variations of consciousness by Marcel Monnier. Participants included representatives of various medical and biological disciplines, psychologists, sociologists, and anthropologists. The difficulties the group experienced in attempting to reach a commonly accepted definition of consciousness are an eloquent example of similar difficulties in many communications of psychologists.—*A. R.*

Burlingham, Dorothy. *Twins: a study of three pairs of identical twins.* New York: International Universities Press, 1953. Pp. x + 92 + 30 charts. $7.50.

This report contains 30 developmental charts with mainly qualitative notations in various areas of behavior. Little discussion is given to the methodological problems involved in such a study. The greatest emphasis is placed upon the twins' relationships to each other and to the social world around them. Attention is called to the active and passive roles that members of the pair assume. The phantasy of having twins is compared with the reality of twinships. Acquired differences are shown to be related to various factors. Identification is indicated as a process which keeps the twins identical in spite of their acquired differences.—*F. McK.*

Davidson,Audrey, & Fay, Judith. *Phantasy in childhood.* New York: Philosophical Library, 1953. Pp. viii + 188. $4.75.

This book discusses the meaning of childhood phantasy from a determinedly authoritarian Freudian point of view. At times the interpretations seem to transcend any reasonable limits of valid logical inference, and it is definitely not a book for the experimentally minded. The case material is abundant and well chosen for interest, however, and if the reader is displeased by the rigidly analytic interpretations placed upon the phantasy material by the authors, he may substitute his own. The theoretical context is that of Melanie Klein and Susan Isaacs, and the authors' approach is expository rather than critical or creative. The volume is well written, presented in an attractive format, and offers pleasant reading.—*W. A. H.*

Garrett, James F. (Ed.) *Psychological aspects of physical disability.* Federal Security Agency, Office of Vocational Rehabilitation, Rehabilitation Service Series No. 210. Washington: U. S. Government Printing Office, 1952. Pp. vii + 195. 45¢.

Psychological Aspects of Physical Disability gives a clear description of the psychological phases of the various types of handicaps. Although 18 experts have written chapters in their special fields, the contributions are so well blended and edited that one has a feeling of a well-integrated whole. There is an excellent chapter on psychiatric aspects of physical disability, and another on the social psychology of adjustment to physical disability. While the book is planned primarily for vocational rehabilitation counselors who are face to face with adult problems, it contains so many practical suggestions that it would be a valuable addition to the library of any psychologist. The book emphasizes the need of reaching the individual, inspiring him to rehabilitate himself.—*B. M. L.*

Garrett, Henry E. *Statistics in psychology and education.* (4th Ed.) New York: Longmans, Green, 1953. Pp. xii + 460. $5.00.

This old reliable text has had only a minor face lifting. There are some added materials on analy-

---

Note.—The reviews were prepared by the Editor and the Associate Editors, who may be identified by their initials.

sis of variance, new problems, several rearrangements, and a new, attractive format.—*L. F. S.*

Harrower, Molly. *Appraising personality.* New York: Norton, 1952. Pp. xvii + 197. $4.00.

This attractive and highly literate presentation of psychological techniques to persons in allied professions is cast in the form of a dialogue between a psychologist and his physician friend. Gently, with case studies and other simple illustrations, the physician is led to the psychologist's use of the Rorschach, the Wechsler-Bellevue, figure drawings, incomplete sentences, and the Szondi. A final section shows the integration of psychological methods in solving five difficult diagnostic problems. The book's shortcoming is that it is almost too good. Persuasive exposition, insufficiently tempered with critical appraisal, can oversell techniques whose real validity is in doubt.—*L. F. S.*

Kluckhohn, Clyde, Murray, Henry A., & Schneider, David M. *Personality: In nature, society, and culture.* (2nd Ed.) New York: Knopf, 1953. Pp. xxv + 701 + xv. Text edition, $5.75.

In the revision of this useful sourcebook, the introductory essay on personality by Murray and Kluckhohn has been completely rewritten and considerably expanded. The essay, developed from the formula "tension—reduction of tension," is itself a 47-page miniature treatise. The readings include 32 of the 39 selections which were in the first edition, together with 13 new ones, two of which are on the uses of literature for psychology.—*L. F. S.*

Moustakas, Clark E. *Children in play therapy.* New York: McGraw-Hill, 1953. Pp. ix + 218. $4.50.

Describing the play therapy program with preschool children at the Merrill-Palmer School of Detroit, the author makes a great point of the difference between nondirective and client-centered therapy on the one hand, and the "child-centered" therapy discussed in this book, on the other. The difference is hard to find in the great amount of verbatim recording presented. He criticizes reflection of feelings as a "repetitious, unsympathetic, static response." His case illustrations are often exactly that. There are brief discussions of preventive play therapy, play therapy with normal children, situational play therapy (adapted from the author's article in the *Journal of Consulting Psychology,* June 1951), and play therapy with disturbed children. About half the volume is devoted to the treatment of one preschool child and her parents. In this case, the mother's presence in the play therapy room, brought about by the child's refusal to leave the mother, and the insight the mother thus accidentally gained, were instrumental in the improvement of behavior and attitudes that followed. This book may be a "key to understanding normal and disturbed emotions" for some laymen. Psychologists will find it superficial.—*M. K.*

Sarason, Seymour B. *Psychological problems in mental deficiency.* (2nd Ed.) New York: Harper, 1953. Pp. x + 402. $5.00.

Sarason's excellent book on mental deficiency has been extended rather than revised. Its usefulness has been increased by three added chapters on practical problems: the interpretation of mental deficiency to parents, criteria and procedures for institutionalization, and problems of professional training.—*L. F. S.*

Skinner, B. F. *Science and human behavior.* New York: Macmillan, 1953. Pp. x + 461. $4.00.

An introductory textbook in psychology which contains not a single table, figure, or explicit account of an experiment is indeed a rarity these days, especially when the author is an outstanding experimentalist. Skinner seems to take the experimental basis of psychology for granted, although he pays considerable attention to its broader foundations as science. The most evident aspect of the author's position is his distaste for intervening variables. "Drive," for example, is not a stimulus, a physiological state, a psychic state, or a state of strength but ". . . a convenient way of referring to the effects of deprivation and satiation and of other operations which alter the probability of behavior. . . ." To control behavior, the drive cannot itself be manipulated directly, but only the conditions of deprivation and satiation. The application of this conceptual framework to complex problems such as the self, anxiety, control, religion, and psychotherapy makes interesting reading. Furthermore, the author knows his literature, his philosophy, and his Freud, and bends them to his own ends. The result is a stimulating and challenging exercise for psychologists, although not all will be convinced. A different evaluation of the volume must be given with respect to its function as an introductory textbook. This book knows all the answers. The student is therefore likely to perceive psychology as a neatly wrapped bundle of truths, and is deprived of the opportunity to see scientific method as an ongoing process by which successive approximations of truth are reached without ever coming to an end. —*L.F.S.*

Tests

Blum, Lucille H., & Fieldsteel, Nina D. *Blum-Fieldsteel Developmental Charts.* Ages 0–6 yrs. Set of 2 charts ($2.50 per 25), with manual, pp. 8. Yonkers, N. Y.: World Book Co., 1952, 1953.

Two well-designed cumulative graphs are used to plot a child's development according to the observations and norms of Gesell and Amatruda. In each chart, the observations are plotted as an age-calibrated scale on the ordinate, with chronological age as the abscissa. The chart of Motor Behavior is concerned mainly with postural and locomotor development from "head erect" to "standing on each foot alternately with eyes closed." The chart of Functional Behavior ranges from "holds objects" to "ties shoelaces." The manual suggests uses of the charts by psychologists, pediatricians, nursery schools, rehabilitation classes, and others, with normal and handicapped children.—*L. F. S.*

Maslow, A. H. *The S-I [Security-Insecurity] Inventory.* College-adult. 1 form. Untimed, (10) min. Inventory ($1. per 10, $7. per 100), with manual, pp. 10, and key. Stanford, Calif.: Stanford Univer. Press, 1952.

Maslow's 75-item questionnaire is the end product of a series of case studies and item analyses over a period of more than 10 years. The outcome is a technique far more refined than most short questionnaires, with good reliability (about .90) and some clinical evidence of validity. Numerous applications are suggested, for research, screening, student personnel services, and counseling. — *L. F. S.*

Books Received

Andrews, T. G. (Ed.) *Méthodes de la psychologie.* Paris: Presses Univer. de France, 1952. 2 vols. Pp. vii + 882. 3.000 fr.

Bernhardt, Karl S. *Practical psychology.* New York: McGraw-Hill, 1953. Pp. xii + 337. $3.75.

Geldenhuys, J. Norval. *The intimate life.* New York: Philosophical Library, 1952. Pp. 96. $2.75.

Halmos, Paul. *Solitude and privacy.* New York: Philosophical Library, 1953. Pp. xvii + 181. $4.75.

Kraft, Victor. *The Vienna circle.* New York: Philosophical Library, 1953. Pp. xii + 209. $3.75.

Ray, Marie Beynon. *The best years of your life.* Boston: Little, Brown, 1952. Pp. xiv + 300. $3.95.

Reiss, Samuel. *The universe of meaning.* New York: Philosophical Library, 1953. Pp. xi + 227. $3.75.

Runes, Dagobert D. *The Soviet impact on society.* New York: Philosophical Library, 1953. Pp. xiii + 202. $3.75.

Tead, Ordway. *Character building and higher education.* New York: Macmillan, 1953. Pp. x + 129. $2.00.

Teicher, Joseph D. *Your child and his problems.* Boston: Little, Brown, 1953. Pp. x + 302. $3.75.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Some Test Correlates of Response to Psychotherapy[1]

Frank Barron
Institute of Personality Assessment and Research
University of California

This paper reports certain personality and intelligence test correlates of changes in psychological morbidity in 33 adult psychoneurotics following six months of psychotherapy in the outpatient service of a state psychiatric clinic. These findings are part of the results of a study of somewhat broader scope, in which psychotherapy was considered as a special case of personal interaction in general, and in which therapeutic relationships and outcomes were rationalized in terms of the personality structures of therapists as well as of patients [1].

The present paper, in addition to being restricted to a report on the test results only of patients, will adopt by choice a somewhat narrow conception of response to psychotherapy. The criterion variable can be best characterized as an answer to the question, "Did the patient improve or not?" where the word "improve" means some fairly general changes in state from bad to good. Examples of such changes are these: (*a*) The patient feels better, is more comfortable, takes more interest in life, and the like; (*b*) important interpersonal relations are straightened out a bit; (*c*) physical symptoms have been relieved or cured; (*d*) important health-tending decisions have been made; (*e*) there has been an increase in insightful remarks and behavior.

The judgment as to whether or not improvement had occurred was made by two expert observers (heads of departments at Langley Porter Clinic) who had not had any part in the conduct of therapy. Each therapist made a formal presentation before these two judges of every case he had handled; prior to the presentation the judges had read all of the material concerning the patient which had been recorded in the clinical chart.

In the instruction to the judges it was emphasized that the crucial variable was *not* general level of functioning of the patient at the conclusion of the therapy, but rather the *change in state* which had occurred between beginning and end of therapy. Further, it was made clear to the judges that part of their function was to evaluate the therapist's involvement in his own account of the therapeutic process, and to weigh that factor in coming to a best estimate as to the degree of change which had actually occurred.

On the basis, then, of two main sources of information (formal presentation of the case by the therapist, and an evaluation from the clinical chart), the expert raters assigned cases, first of all, into two main categories, those who had shown definite improvement and those who had failed to improve or who had improved only slightly. They were in close agreement on this crude categorization, differing on only 2 of the 33 cases which were studied and rated. In addition, they assigned ratings on a 100-point scale to each case, so that the total sample was ranked in terms of the variable "Improvement." The correlation between the two independent sets of ratings was .91. The two disagreements were resolved by averaging the scale ratings. The final classification included 17 patients in the Improved group, 16 in the Unimproved.

Some Illustrations of Criterion Behavior

It is of some interest to examine the kinds of incidents and outcomes which the judges considered indicative of improvement or lack of it. Here is a partial list of such incidents and outcomes, abstracted from the clinical charts of these patients:

---

[1] This study was conducted for the most part at Langley Porter Clinic, San Francisco. It was completed at the Institute of Personality Assessment and Research, and was supported financially by the grant of the Rockefeller Foundation to that Institute. The author is indebted to Robert E. Harris, director of the psychology service at Langley Porter, for many valuable suggestions which contributed to the research.

Improvement:

1. A woman patient who had been frigid through four years of marriage which terminated in divorce now established an intimate relationship with another man in which for the first time she experienced orgasm in sexual intercourse.
2. A man who entered therapy in a very depressed, anxious, and physically upset state, and whose troubles centered on his relations with his foreman on a construction job (a relationship in which he was outwardly submissive and cooperative but inwardly enraged), finally learned to "stand up to" the foreman and express his feelings. There was a clear advance in his feelings of independence and self-esteem, and towards the conclusion of the therapy the patient left his former job and started a business of his own.
3. A woman with menstrual difficulties related to unconscious hostility towards her husband because of his disregard of her own wishes in sexual intercourse became conscious both of her anger and her retaliative tactics, with consequent disappearance of the menstrual difficulties and a more candid and unmartyred relationship to her husband.
4. A woman who had been recently divorced after an extremely traumatizing marriage and who entered therapy in a state of tremendous agitation and anxiety, so incapacitating that she could not sit in the lobby of the clinic or cross a street or look anyone in the face, gradually lost her phobias, experienced a great reduction in anxiety, and became able to associate comfortably with other people.
5. A man suffering from severe gastric disturbances which he considered unrelated to any emotional problems entered therapy upon the strongly worded advice of the internist. His dissociation of affect from the death of his mother and subsequent disturbing childhood years in an orphanage was gradually broken down and he reexperienced his own sense of forsakenness and finally came to terms with it, crucially aided by the support and understanding of the woman therapist.

Lack of improvement:

1. A very dogmatic and dominating man who entered therapy at the insistence of his wife (who complained of his overbearingness with her and their three children) proved too overbearing for the young woman therapist as well. The therapist was unable to handle the negative countertransference which developed, and the therapeutic interaction became simply a battle, with consequent lack of improvement in the patient.
2. A woman of forty, who had never been married and who suffered from continual headaches and disturbing sexual impulses which she attempted to suppress, found herself obliged to terminate therapy for some good reason precisely after the hour at which the therapist had told her he would give her a physical examination on the occasion of their next appointment.
3. A man with a history of homosexuality attempted to seduce his male therapist, who responded with anger as well as some anxiety. The patient did not return.
4. A psychopathic young man who had a long history of rebelliousness, brushes with the law, and employment instability entered therapy because, he said, he happened to be passing by the clinic and wondered what it would be like to be a patient. He liked it for a while, especially as he happened to be assigned to an attractive woman therapist, but after a few months he hitchhiked off in a southerly direction, leaving a chatty farewell note for the doctor in which he addressed her by her first name.
5. An engineer of extremely masculine appearance and manners complained of a "psychic malfunction" and wanted "a few treatments to get at the nuclear material and get the thing cleared up." His therapist insisted on taking a nondianetic, unhurried approach to the matter, which annoyed the patient, who nevertheless remained in therapy in the expectation that something would eventually happen.

One might summarize the kinds of outcomes in the Improved group as follows: (*a*) The patient generally reported a feeling of well-being at the conclusion of therapy, in contrast to depression and anxiety at its start; (*b*) specific symptoms, such as headaches, frigidity, or impotence, gastric disturbances, menstrual difficulties, skin disorders, etc., tended to be relieved or totally cleared up; (*c*) in some cases there were significant changes in the direction of more mature interpersonal relations, especially with parents, parent-substitutes, or spouses.

The failures in psychotherapy could often be traced to the inability of the therapist to handle some particularly difficult problem. Perhaps with more experienced psychotherapists some of the patients who did not improve would have made some progress; however, there seemed little doubt that the cases which were marked down as therapeutic failures were basically more difficult problems. As we shall see later from the test results, it was generally the more disturbed individuals who did not improve.

Test Correlates of Improvement

The test battery given these patients before psychotherapy began consisted of a shortened form of the Wechsler-Bellevue Intelligence Scale, the Minnesota Multiphasic Personality Inventory, The Rorschach Psycho-diagnostic and the Ethnocentrism Scale (in Form 60 of the University of California Public Opinion Study scales). The patients themselves were all receiving psychotherapy for the first time, and they began the course of treatment just at the time that their therapists were starting a six-month period of service in the outpatient clinic. The therapists were psychiatric residents who had had little experience in intensive psychotherapy.

Most of the patients had one hour of treatment a week for six months. Of the 33 patients, 12 were men and 21 were women. They ranged in age from 20 to 45. All but 3 of them were high school graduates, but only 2 were college graduates. As a group they could probably be characterized as lower-middle-class socioeconomically. Two-thirds of them were gainfully employed.

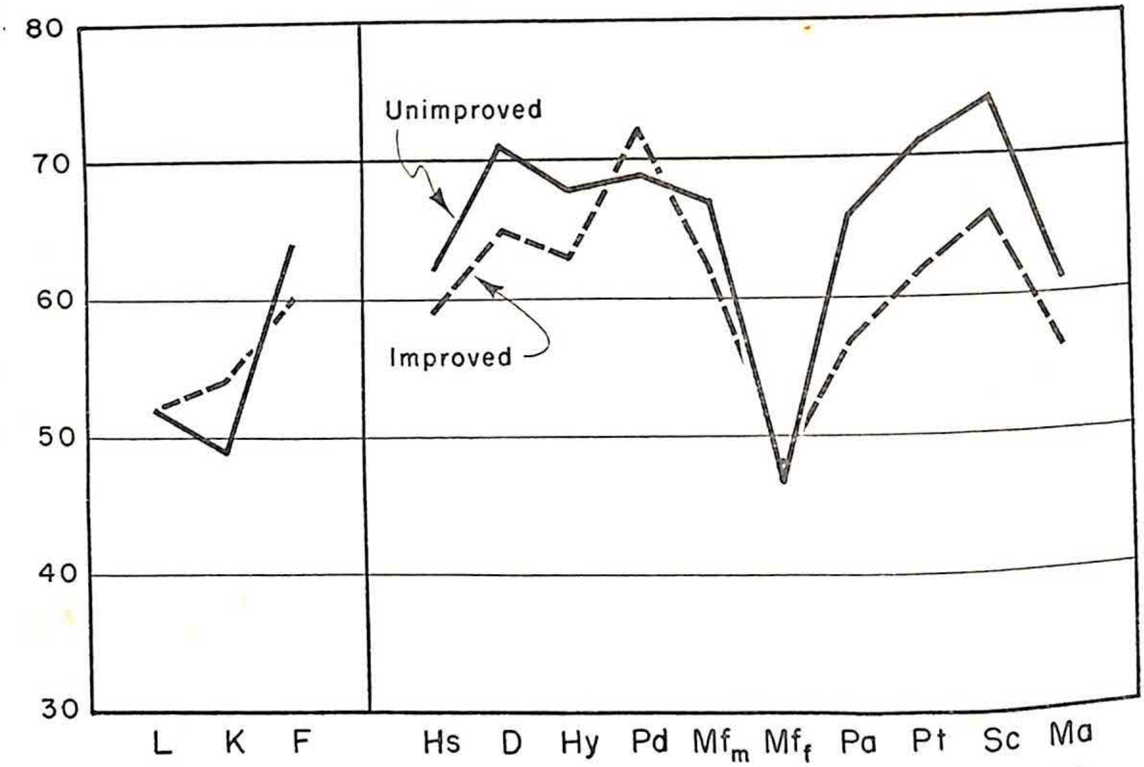
All of the testing was done by the experimenter. Conventional controls were exercised to prevent contamination in the prediction of outcome from the tests. The controls were as follows: (*a*) Only the experimenter knew the test results. The psychotherapists were informed of this condition before the experiment began, and were told that they would be given information concerning the testing only if in their judgment it was essential to the diagnostic formulation and the handling of the case. In three cases such information was requested by the therapist during the course of the therapy. The information given was not comprehensive, but was in response to some specific point. These cases were retained in the sample in spite of their being exceptions to the general rule, for their exclusion would be as likely a source of contamination as their retention. (*b*) The judges of outcome had no knowledge of the test results. (*c*) The test protocols were identified by a code number rather than by name, so that clinicians who attempted to predict outcome had no knowledge of the patient beyond that afforded by the test results themselves.

The design of the present study, so far as prediction of outcome is concerned, is a replication of some earlier work by Harris and Christiansen [5], who used in an almost identical clinical setting a similar criterion of outcome of therapy and the same clinical tests, except that they did not employ the Ethnocentrism scale. It is therefore in some respects a cross-validation of their findings, or at least it offers further evidence on some of the questions they discussed. Their results will therefore be compared with the present findings whenever such comparisons seem relevant.

The Wechsler-Bellevue Intelligence Test. An abbreviated form of the Wechsler scale, consisting of the Comprehension, Similarities, Block Design, and Digit Symbol subtests, was used. Harris and Christiansen, using the same subtests, had found no significant relationship between Improvement and Intelligence. In this sample, however, the two variables were associated positively, the Pearsonian r being .46, which is significant at the .01 level. The mean of the Improved group was 122, that of the Unimproved group was 112.

The mean IQ of the total sample was 117, as compared with a mean of 108 in the Harris-Christiansen study. An important difference between the two samples so far as method of selection is concerned should be noted: The Harris-Christiansen sample consisted of already hospitalized persons who were given psychotherapy because of delayed recovery from physical disease, surgery, or accident, whereas almost all of the patients in the present study had elected on their own initiative to seek psychotherapy at the clinic, and in their waking hours went about the business of life in an upright position. This difference might well account for the IQ difference between the samples, as well as for the fact that these self-referred patients were a good standard deviation above the general population mean in intelligence. Simply being aware of the fact that psychotherapy is to be had and that it makes sense to seek it when you are in personal difficulties is probably related positively to general intelligence and cultural sophistication.

It is also to be expected on theoretical grounds that greater effectiveness of cortical functioning (which presumably reveals itself on the Wechsler-Bellevue) should be associated with a factor of modifiability in personality structure. Intelligence certainly involves the ability to cognize relationships adequately, including emotional relationships, and to correct



	L	K	F	Hs	D	Hy	Pd	Mf_m	Mf_f	Pa	Pt	Sc	Ma
Improved	52	54	60	59	65	63	72	62	48	57	62	66	56
Unimproved	52	49	64	62	71	68	69	67	47	66	71	74	61

Fig. 1. Group MMPI profiles of the Improved and Unimproved groups.

one's cognitions on the basis of new evidence. It would seem that in this sample, at any rate, the more intelligent patients were better able to use the psychotherapeutic relationship to induce desired personality changes.

The Minnesota Multiphasic Personality Inventory. The MMPI analysis may be divided into three main sections: (*a*) group differences between the Improved and Unimproved cases on MMPI scales; (*b*) prediction of outcome from MMPI profiles by clinical psychologists skilled in the use of the test, as well as by certain rational indices or rules for mechanical sorting of the profiles; (*c*) development and cross-validation of a prediction scale. Enterprises 1 and 2 will be reported in this paper; the development and cross-validation of a prediction scale will be reported separately, however, as it involved work with a number of other samples and raised some test-specific questions which are not properly within the scope of this paper [2].

Group comparisons on the usual diagnostic and validity scales are shown in Figure 1. Only on the Paranoia scale was there a significant difference between the means of the Improved and Unimproved groups (t of 2.6, $p < .02$). The Unimproved group profile is consistently higher on almost all scales, however, and the peak is at Schizophrenia, with the average T scores for Psychasthenia and Depression also being greater than 70. The sorts of individuals who earn profiles like this are usually schizoid; in *An Atlas for the Clinical Use of the MMPI* [6], most of the 8724 profiles carry a diagnosis of reactive depression with underlying schizoid trends (pp. 668-669). This fits well with the observation by Harris and Christiansen that poor prognosis is generally associated with what they call "subclinical psychotic trends," referring to a more severe ego dysfunction than first clinical impression seems to indicate. Depression and anxiety are salient features of the clinical picture in almost all patients who seek psychotherapy in the outpatient clinic; but the patients who fail to improve are those in whom paranoid and schizoid features underlie the psychoneurotic symptoms. A reasonable guess would be that their personal difficulties are more chronic and characterologically based, in contrast to the more acute and situation-linked problems of the patients who improve.

The 33 individual MMPI profiles were also given to clinical psychologists for prognostic evaluations. Eight clinicians independently attempted to predict outcome from the profiles, having no other information concerning the patient beyond age and sex. Each clinician sorted the 33 records into two groups, predicting either that the case would be improved or that it would be unimproved. Of 264 (8×33) such classifications, 164 (62.12%) were correct. The sum of the eight values of chi square was 20.46, which for 8 *df* yields a p less than .01. The three best sorters had a pooled accuracy percentage of 69.7%. All of the sorters were more often correct than in error, but it should be pointed out that an average accuracy of 62% when 50% would be expected by chance is a fairly modest achievement in clinical prediction. Even the best of the MMPI analysts would be well advised to be tentative in their prognostic formulations.

Rational indices and rules for mechanical sorting produce substantially the same accuracy of prediction as do sortings by clinicians; actually, such indices are slightly more effective. Two fairly simple and obvious mechanical predictors were put to the test. The first of these employed the rule that for any profile with all scores within the normal range (i.e., greater than 30, less than 70) Improvement should be predicted. Ten profiles met this requirement, and 9 of these cases had in fact improved. Twenty-four of 33 profiles (73%) were thus classified correctly (chi square of 8.1, $p<.01$). Exceptions were then allowed in the general rule: e.g., profiles with only D, or with only D and *Pt*, greater than 70 would be called Improved. Both exceptions led to an accuracy of prediction between 75 and 80%.

In addition to these sorting rules, a simple index of subclinical psychotic trends was defined by summing the scores of four scales (*F*, *Pa*, *Sc*, and *Ma*) usually found elevated in psychosis. This index correlated .40 with outcome as it had been rated on a 100-point scale.

It is of some interest that the same indices which predicted response to psychotherapy worked equally well in another setting, with a sample apparently quite different from the patient sample at this clinic. In a study of personal effectiveness in male graduate students in their final year of work for the doctorate, ratings were obtained from faculty members on certain dimensions such as originality and personal soundness. The averaged faculty ratings on personal soundness for 40 such students correlated—.4 with the index defined by the sum of *F*, *Pa*, *Sc*, and *Ma*, and had a contingency of .52 with a mechanical sorting of the MMPI profiles using the rule that all profiles having no clinical scale elevation (except *D*) of 70 or above should be called Sound, while all others should be called Unsound.

The "personally sound" graduate students were different from the "personally unsound" students in the same way that patients who profited considerably from psychotherapy differed from patients who profited little or not at all. As one would expect, however, there was a difference in absolute level of adjustment; the distribution of scores on the Prognosis for Psychotherapy scale [2] showed a clear progression in group means from the Unimproved patients through the Improved patients and the Unsound graduate students to the Sound graduate students.

What these relationships indicate, it would seem, is that the patients who are most likely to get well are those who are not very sick in the first place. Another way of putting this is to say that patients who are more integrated to begin with are better able to use the psychotherapeutic relationship to solve whatever problems brought them into therapy. A corollary of this is that the potential gains from therapy should increase as therapy progresses, at least up to the point where the critical problem for the patient is to become genuinely independent of the therapist. By that time, of course, the therapeutic process has become internalized, and personality problems have been brought "into the ego," to be dealt with rationally and objectively.

The Rorschach Psychodiagnostic. The Rorschach analysis was carried out according to the same general design as the MMPI analysis, with the exception that no prognostic scale was developed. Average profiles were determined for the Improved and Unimproved groups, using the Klopfer-Kelly conventions for the scoring and nomenclature of determinants. Individual scores on the Harris-Christiansen Rorschach Prognostic Index [5] were computed, and the relationship of Outcome to that index was calculated. Finally, the test protocols, including both content of responses and profiles of determinants, were given to clinicians for sorting.

All of Rorschachs were administered by the experimenter, with all of the verbalizations and relevant behavior of the patient being recorded. The tests were scored independently by the experimenter and by one other clinical psychologist who had had several years of experience both in the scoring and the interpretation of the Rorschach. The experimenter himself had had approximately three years of clinical experience with the test, in addition to a graduate course in its scoring and interpretation. The two independent scorings showed a high degree of agreement; discrepancies were reconciled in a conference on review of the scoring. The Improved and Unimproved groups did not differ significantly on any Rorschach determinant, nor on any of the important ratios, such as *W*, *D*, *Dd*, plus *S*, *M*: sum *C*, and so on. This finding is consistent with Rorschach results in the gradute student sample previously referred to; in that study no significant relationship was found between any Rorschach determinant or combination of determinants and the criterion variable, Personal Soundness.

The Harris-Christiansen Prognostic Index correlated .00 with Improvement. As Cronbach [3] has pointed out, most empirical Rorschach indices fail to hold up under cross-validation, so that this is not a surprising finding.

Somewhat contrary to the expectation of most persons who have worked with the test, however, is the finding that experienced Rorschach interpreters were not able to predict outcome of therapy on the basis of the Rorschach evaluation. Four clinicians, all of whom were highly trained and experienced, both in the use of the test and in general clinical practice, attempted the sorting task. They had at hand all of the responses of the patients, as well as a summary of the scoring in terms of the usual ratios and profile of determinants. Only one of the four sortings approached significance, and that was in the negative direction; that is, the sorter was more often wrong than right in predicting outcome.

The Ethnocentrism scale. Certain theoretical considerations had led to the inclusion of the Ethnocentrism scale in the test battery. Levinson [7] had found, in a study at this same clinic, that ethnocentric patients showed a personality syndrome of a sort which would be resistant to change through psychotherapy. High-scoring patients were described by her as having the following characteristics: "rigid, constricted personalities . . . stereotyped, conventionalized thinking . . . undifferentiated egos . . . narrow range of experience, emotionally and intellectually . . . weak interpersonal relationships, expressed in terms of dominance-submission." The low-scoring patients, on the other hand, were marked by their "desire to be loved and fear of being rejected . . . expressed directly in interpersonal relations and in a very personalized way . . . the frequency with which low scorers discuss their relationships to others is striking."

These considerations suggested the hypothesis that patients who scored high on the Ethnocentrism scale would be less likely to enter deeply into the sort of personal interaction which occurs in psychotherapy. From the very beginning of therapy they would find it necessary to narrate the comings and goings of their somatic complaints, and to minimize the importance of their psychological problems as the cause of their somatic disturbances. According to this notion, ethnocentric patients would tend characteristically to isolate affect and to avoid psychological self-examination; they would be strangers to their own inner life.

While the psychological formulation itself was not checked on in any detail in the present study, the hypothesis that it led to (i.e., that Ethnocentrism would be negatively related to Improvement) was confirmed. The correlation of the *E* scale with Improvement ratings was —.64, making it the best single predictor employed in this study.

Several reservations, however, should be noted. In the first place, scores on the *E* scale are negatively related to intelligence, which can be presumed to be a somewhat more basic psychological variable than ethnocentrism. As Gough [4] has pointed out, many of the items in the *E* scale are written in such an extreme fashion that a person of some intelligence and education could not agree with them, even though the sentiment expressed might win his approval if the form of its statement were less clearly irrational. Since Improvement is correlated positively with Intelligence in this sample (.46), one would expect a negative relationship between Improvement and Ethnocentrism on the basis simply of their common relationship to Intelligence.

When Intelligence is partialled out, however, some relationship remains. The coefficient of partial correlation between Ethnocentrism and Improvement, with Intelligence held constant, is—.34. Partialling intelligence out in this manner is not strictly correct, of course, except in a mechanical way; actually, some underlying emotional-cognitive variable such as rigidity, which is complexly related both to intelligence and to ethnocentrism, may be accounting for a large part of the correlation of both of those measures with the capacity for psychological change as shown by improvement in therapy.

There is one other reservation to be noted. Four of the psychotherapists in this study were Jewish, a fact which might have affected both the ethnocentric patient's attitude towards them and their attitude towards the ethnocentric patient. In several of the therapeutic interactions, such patient attitudes came out into the open, and were dealt with therapeutically; and in at least one case the therapist admitted candidly to the psychologist that his own response to the patient's anti-Semitism was disruptive of the relationship.

In spite of these reservations, however, the relationship between ethnocentrism and capacity for psychological change through psychotherapy is a striking one, and if confirmed would be of some theoretical importance. It may, indeed, have implications for social communities as well as individuals, and would suggest that the social organization which fosters prejudice might well be forfeiting the kind of emotional flexibility which is necessary if in time of crisis it is to cure its own ills.

Summary

Change in psychological morbidity as a consequence of six months of psychotherapy was studied in a sample of 33 adult psychoneurotics who were treated in the outpatient service of a state psychiatric clinic. Changes for the better might be summarized as follows: (*a*) The patient generally reported a feeling of well-being at the conclusion of therapy, in contrast to depression and anxiety at its start; (*b*) various psychosomatic symptoms tended to be relieved or cured; (*c*) significant interpersonal relations were established on a more mature basis.

Patients were rated on the degree of their improvement by two skilled judges who evaluated both the interview-by-interview account of the therapy in the clinical chart and the therapist's report in conference on the course and outcome of the therapy. Ratings of degree of change for the better were then correlated with various psychological tests.

The chief test correlates are these: (*a*) Intelligence is positively associated with change; (*b*) level of integration at the beginning of therapy is positively associated with change (i.e., those who are better off to begin with are likely to improve the most); (*c*) ethnocentrism is negatively related to change.

Received December 11, 1952.

References

1. Barron, F. Psychotherapy as a special case of personal interaction: prediction of its course. Unpublished doctor's dissertation, Univer. of California, 1950.
2. Barron, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.,* in press.
3. Cronbach, L. Statistical methods applied to Rorschach scores: a review. *Psychol. Bull.,* 1949, **46,** 393–429.
4. Gough, H. G. Studies of social intolerance: I. Some psychological and sociological correlates of anti-Semitism. *J. soc. Psychol.,* 1951, **33,** 237-246.
5. Harris, R. E., & Christiansen, C. Prediction of response to brief psychotherapy. *J. Psychol.,* 1946, **21,** 269-284.
6. Hathaway, S. R., & Meehl, P. E. *An atlas for the clinical use of the MMPI.* Minneapolis, Minn.: Univer. of Minnesota Press, 1951.
7. Levinson, Maria H. Psychological ill-health in relation to potential fascism: a study of psychiatric clinic patients. In T. W. Adorno, *et al.,* (Eds.) *The authoritarian personality.* New York: Harper, 1950. Pp. 891-970.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

A Nonintellectual Intelligence Test [1, 2]

Harrison G. Gough
University of California, Berkeley

The title of this paper may be something of a misnomer, for it is not at all certain that the testing instrument which is going to be presented is, actually, a measure of intellect. Nevertheless, it does correlate significantly with conventional tests of intelligence, and seems to reveal other properties which suggest that something along the line of intellectual ability, or perhaps more precisely intellectual efficiency, is being appraised. The other part of the title—the "nonintellectual" part—has not been the cause of any misgivings, because, as will be seen shortly, the test instrument does not call for any problem solving, reasoning, abstract thinking, or any of the other mental functions which are traditionally regarded as the key factors in defining and measuring intelligence.

The two immediate aims of this project were first, to determine whether a criterion-specific set of personality inventory items could be assembled which would correlate significantly with accepted measures of intellect, and second, if this was successful, to determine whether the individual items could be combined in a scale which would have anything approaching practical significance and value.

The second objective is added, because, as we know for instance from work on the Rorschach [1], a series of variables can easily be identified which will correlate significantly with intellect, but which will seldom rise above the level of + .20 and + .25 into a region of more dependable utility when pooled in a composite.

Procedures

The first step in the present project was to write a series of criterion-specific inventory items, and also to select appropriate items from standard sources. An example of a spe-

Table 1

Item Statistics for Four Items from the *Ie* Scale

Item	Percent saying "True"					
	Higher scores on intellectual ability measures			Lower scores on intellectual ability measures		
	School A (*N* 28)	Schools B and C (*N* 21)	School D (*N* 48)	School A (*N* 28)	Schools B and C (*N* 21)	School D (*N* 48)
1. It is more important that a father be kind than that he be successful.	79	95	92	71	76	69
2. Parents are much too easy on their children nowadays.	36	43	46	50	71	67
3. I am not afraid of picking up a disease or germs from door knobs.	86	81	85	79	67	67
4. Most people make friends because friends are likely to be useful to them.	21	19	29	46	71	77

---

[1] This paper is one of a series of studies concerned with the development of practical assessment devices for the identification of positive and favorable aspects of personality and temperament. The project is being carried out under a research grant from The National Institute of Mental Health, National Institutes of Health, U. S. Public Health Service.

[2] The present paper is a revision and extension of a preliminary report given at Fresno, California, April 26, 1952 at the annual meeting of the Western Psychological Association.

Table 2
Summary Statistics for the First (76-item) Intellectual Efficiency (*Ie*) Scale in the Original Samples

Sample	*N*	*M*	*SD*	r*
School A	135	52.44	8.51	.47
School B	35	51.34	9.20	.68
School C	66	51.41	9.41	.64
School D	205	51.93	9.51	.54

* Intercorrelations between the Ie scale and a conventional IQ test in each of the schools.

Table 3
Summary Statistics for the First *Ie* Scale in Six Cross-Validating Samples *

Sample	*N*	*M*	*SD*	r†
I. For the abbreviated (59-item) scale.				
1. St. Cloud ninth grade	82	38.09	7.46	.53
2. Minneapolis ninth grade	190	39.44	8.24	.36
3. St. Cloud twelfth grade, 1947	231	39.55	7.50	.52
4. St. Cloud twelfth grade, 1948	271	40.30	7.70	.47
5. University of California advanced graduate students	40	48.58	4.02	.42
II. For the full (76-item) scale				
1. University of California first-year graduate students in psychology	30	63.20	5.12	.44

* The first five samples were scored only for the 59 items taken from the MMPI. The psychology sample took the full scale.

† For the two graduate student samples, the criterion test was the Miller Analogies Test, form g. For the other samples, conventional IQ measures were used.

cially written item would be, "I get sort of annoyed with writers who go out of their way to use strange and unusual words," and of an item from a standard questionnaire, in this case the MMPI [5], "I gossip a little at times."

These items were then reproduced in a booklet and administered to the senior students in four Minnesota high schools.[3] In each school, the responses of students in the upper 25 per cent of the range on a conventional intelligence test were compared with those of students in the lower quartile.[4]

Table 1 provides an example of the way in which the analysis was conducted. Items which showed consistent and significant differences in all Hi-Lo comparisons were retained for the scale. Altogether, 76 items were identified which met this criterion.

These 76 items were scored on the original samples with the results indicated in Table 2. The next step was to cross-validate this scale. Table 3 presents the results observed in six cross-validating samples.[5] In the first five, only MMPI [5] items were available, but the sixth sample took the full 76-item scale. All of the correlations are clearly significant, statistically, and all save one are high enough to be of some practical value. An interesting finding here was the degree to which the scale retained its validity when checked in the graduate student samples where one would encounter rather severe attenuation of the criterion measure.[6]

The individual items were also checked in the two graduate student samples, and it was discovered that 24 of them were contributing only slightly, if at all, to the correlations with the criterion tests. It was decided, accordingly, to eliminate them and to reduce the scale to 52 items.[7] This 52-item scale is the one which is currently being used, and it is also the version included in the California Psychological Inventory.[8,9] These 52 items, and the scored responses, are listed below.

---

[3] The four communities were: Bemidji, Dawson, Redwood Falls, and St. Cloud. The writer wishes to thank the following persons for making this testing possible: Mr. R. B. Clay, Mr. C. H. Dahlin, Mr. H. B. Gough, Mr. R. Gray, Mr. J. W. Smith, Mr. R. H. Sorenson, Mr. G. Ramseth, and Mr. R. H. Witt.

[4] Because of the small number of cases available, schools A and B were combined into a single sample for this analysis.

[5] The first four samples were made available through the kindness of Mr. M. J. Lunemann, Dr. Starke Hathaway, Mr. F. J. Herda, and Mr. H. B. Gough. The graduate student samples were seen at the University of California Institute of Personality Assessment and Research, Dr. D. W. MacKinnon, Director, as part of a large-scale study of personal and social effectiveness. The Institute projects are being carried out under a grant from the Rockefeller Foundation.

[6] The preliminary 59-item MMPI version was tried in the Michigan clinical psychology assessment program [6]. The abbreviated *Ie* scale yielded significant correlations with the following criterion variables: diagnostic competence (.21); individual psychotherapy (.32); professional interpersonal relations (.28); over-all clinical competence (.23); preference for hiring (.24); and liking (.23).

The correlations for professional interpersonal relations and for individual psychotherapy were both the second highest obtained among all the objective scales utilized in the Michigan study (including the Miller Analogies Test, 6 Primary Mental Abilities test scales, 14 Guilford-Martin scales, 10 conventional MMPI scales, 59 scales on the Strong Vocational Interest Blank, 9 Kuder Preference Inventory scales, and 6 scales from the Allport-Vernon Study of Values).

1. I have had very peculiar and strange experiences. (False) 2. It often seems that my life has no meaning. (False) 3. When someone does me a wrong, I feel I should pay him back if I can, just for the principle of the thing. (False) 4. Any job is all right with me, so long as it pays well. (False) 5. People often expect too much of me. (False) 6. I get sort of annoyed with writers who go out of their way to use strange and unusual words. (False) 7. Success is a matter of will power. (False)

8. When in a group of people I have trouble thinking of the right things to talk about. (False) 9. It is more important that a father be kind than that he be successful. (True) 10. I was a slow learner in school. (False) 11. I must admit I have no great desire to learn new things. (False) 12. My skin seems to be unusually sensitive to touch. (False) 13. Parents are much too easy on their children nowadays. (False) 14. I read at least ten books a year. (True)

15. I have the wanderlust and am never happy unless I am roaming or traveling about. (False) 16. My parents have often disapproved of my friends. (False) 17. Teachers often expect too much work from the students. (False) 18. I am not afraid of picking up a disease or germs from door knobs. (True) 19. I am quite a fast reader. (True) 20. I have had more than my share of things to worry about. (False) 21. I feel like giving up quickly when things go wrong. (False)

22. In a group of people I would not be embarrassed to be called upon to start a discussion or give an opinion about something I know well. (True) 23. I like to read about history. (True) 24. In school I found it very hard to talk before the class. (False) 25. I dread the thought of an earthquake. (False) 26. I like science. (True) 27. I am bothered by people outside, on street cars, in stores, etc. watching me. (False) 28. I have never been in trouble because of my sex behavior. (True)

29. I often get feelings like crawling, burning, tingling, or "going to sleep" in different parts of my body. (False) 30. I seem to be about as capable and smart as most others around me. (True) 31. I like poetry. (True) 32. I am quite often not in on the gossip and talk of the group I belong to. (False) 33. A windstorm terrifies me. (False) 34. I enjoy a race or game better when I bet on it. (False) 35. Most people make friends because friends are likely to be useful to them. (False)

36. I have never seen a vision. (True) 37. The future seems hopeless to me. (False) 38. I have had no difficulty starting or holding my urine. (True) 39. I often feel as if the world was just passing me by. (False) 40. I get pretty discouraged with the law when a smart lawyer gets a criminal free. (True) 41. The only interesting part of the newspaper is the "funnies." (False) 42. I gossip a little at times. (True)

43. At times I have been so entertained by the cleverness of a crook that I have hoped he would get by with it. (True) 44. I have had attacks in which I could not control my movements or speech, but in which I knew what was going on around me. (False) 45. I have had no difficulty in starting or holding my bowel movement. (True) 46. I like to read about science. (True) 47. I do not read every editorial in the newspaper every day. (True) 48. If people had not had it in for me, I would have been much more successful. (False) 49. I daydream very little. (False)

50. I have often been frightened in the middle of the night. (False) 51. I work under a great deal of tension. (False) 52. I seldom worry about my health. (True)

A review of these items is quite interesting in highlighting some of the personological correlates of general intelligence. The best method of extracting the immediate implications of the items is simply to read the list carefully, but as an aid in forming rational groupings and clusters the following synopsis is provided:

1. Self-confidence and self-assurance, freedom from unsubstantiated fears and apprehensions.

---

[7] A table listing each item, the percentage in each sample responding "True," and the significance tests has been deposited with the American Documentation Institute. Order document No. 3972 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting $1.25 for 35-mm. microfilm or $1.25 for 6- by 8-in. photocopies.

[8] The California Psychological Inventory, copyright 1951. This inventory is designed to assess a number of social and interactional variables. Its emphasis is not on medical, psychopathological, or maladaptive characteristics, but, rather, on more favorable and positive aspects of individual functioning. It contains 19 scales, including measures for social responsibility, flexibility, academic motivation, tolerance, etc. See, for example, references [3] and [4].

[9] Thirty-nine of the 52 items were taken from the MMPI (by permission of the copyright owners), and the remaining 13 items were original. The 39 MMPI items, and their scorings, are as follows (numbers given are for the MMPI group form booklet): 13F, 28F, 33F, 35F, 36T, 37T, 60T, 62F, 63T, 78T, 116F, 122T, 146F, 194F, 198F, 221T, 224F, 225T, 256F, 260F, 267F, 277T, 280F, 304F, 338F, 392F, 448F, 455F, 462T, 464T, 487F, 492F, 521T, 524T, 526F, 541F, 546T, 552T, 559F.

2. Effective social technique and adjustment, sense of social acceptability without dependence on others; not suspicious, touchy, or overly sensitive.
3. Good physiological functioning, absence of minor, debilitating symptoms and complaints.
4. Liking, and respect, for intellectual pursuits; wide range of interests.

Three additional samples were then obtained in order to cross-validate the 52-item revised scale.[10] Table 4 presents the findings. The validity coefficients are about what they were with the first six cross-validating samples, so it appears that the shortening did not weaken the scale's effectiveness.

Table 4
Summary Statistics for the Revised (52-item) *Ie* Scale in three Cross-validating Samples

Sample	*N*	*M*	*SD*	*r**
1. Rock Island, Illinois high school	461	33.90	7.07	.49
2. Wellsville, New York, high school	230	34.08	6.39	.50
3. University of California undergraduate psychology students	43	40.05	5.16	.42

* Intercorrelation between the Ie scale and a conventional measure of intelligence in each sample.

Final high school grade averages were also available for the Rock Island sample. For the females, the correlation between grades and tested IQ was .56, and between grades and the *Ie* scale .59. For males the correlation between grades and IQ was .51 and between grades and the *Ie* scale .31.

In order to provide an over-all picture of the summary statistics at this point, the means and standard deviations for each sample have been proportionately equated to the 52-item test version with the results indicated in Table 5. In this table one can observe a consistent rise in mean scores from junior high school groups, through senior high school and college undergraduate levels, to a final culmination in the graduate student samples.

Adjective Check List

The last data to be presented on this scale have to do with the "social-stimulus values" of persons scoring high and low on it. Forty graduate students seen in an intensive 3-day assessment at the University of California Institute of Personality Assessment and Research were given the scale. Staff judges completed Gough Adjective Check Lists [2] for each as-

Table 5
Ie Scale Summary Statistics for All Samples*

Samples	*N*	*M*	*SD*	*r*	*SE*
I. Original samples					
1. Bemidji twelfth grade	135	35.87	5.82	.47	.09
2. Dawson twelfth grade	35	35.12	6.29	.68	.17
3. Redwood Falls twelfth grade	66	35.16	6.44	.64	.12
4. St. Cloud twelfth grade, 1949	205	35.52	6.50	.54	.07
II. Cross-validating samples					
1. St. Cloud ninth grade	82	31.80	6.52	.53	.11
2. Minneapolis ninth grade	190	34.74	7.26	.36	.07
3. Rock Island high school	461	33.90	7.07	.49	.05
4. St. Cloud twelfth grade, 1947	231	34.84	6.61	.52	.07
5. St. Cloud twelfth grade, 1948	271	35.50	6.78	.47	.06
6. Wellsville high school	230	34.08	6.39	.50	.07
7. University of California undergraduate psychology students	43	40.05	5.16	.42	.15
8. University of California first-year graduate students in psychology	30	43.23	3.50	.44	.19
9. University of California advanced graduate students	40	42.80	3.54	.42	.16

* Scores for samples not taking the final 52-item scale were proportionately corrected to equate them to this standard.

---

[10] The Rock Island, Illinois, sample was obtained through the courtesy of Mr. Charles O. Austin, Jr.; the Wellsville sample through the kindness of Dr. H. D. Babcock; and the University sample through the courtesy of Dr. David Krech and Mrs. Ann Miller.

sessee, and then these observer-reports were composited. Adjectives differentiating significantly between the composite reports on the 10 highest and 10 lowest scorers on the *Ie* scale are given below:

1. Adjectives checked more frequently about higher scorers.

adaptable
appreciative
capable
clear-thinking
conscientious
considerate
cooperative
dependable
energetic
enterprising
helpful
honest
industrious
interests wide
kind
logical
obliging
persistent
planful
pleasant
reasonable
realistic
reliable
responsible
sincere
stable
thorough

2. Adjectives checked more frequently about lower scorers.

aloof
cynical
defensive
deliberate
effeminate
headstrong
impatient
individualistic
moody
resentful
sarcastic
sharp-witted

The general trend of adjectives ascribed is quite clear. High scorers are described as capable, clear-thinking, dependable, logical, realistic, etc., and low scorers as cynical, headstrong, impatient, and resentful.

The adjective findings lend support to the notion that the *Ie* scale is measuring something like personal or intellectual efficiency, with their emphasis on characteristics such as industriousness, planfulness, reliability, and thoroughness.

Summary

A study was made of the relationship between intellectual measures and a series of opinion, attitude, and self-description items, believed, on a priori and theoretical grounds, to covary with intellect. Fifty-two items were found to possess empirical validity. These items were assembled in a scale which yielded median validity coefficients of approximately .58 in four original samples, and of .47 in nine check samples.

Adjectival descriptions of subjects scoring high on the scale tended to stress factors of dependability, intellectual clarity, persistence, and planfulness. The decision was made to call the scale a measure of "intellectual efficiency" in order to reflect these properties.

Received December 3, 1952.

References

1. Altus, W. D., & Thompson, Grace. The Rorschach as a measure of intelligence. *J. consult. Psychol.*, 1949, **13**, 341-347.
2. Gough, H. G. *Predicting success in graduate training.* A progress report. Berkeley: Univer. of California Institute of Personality Assessment and Research, 1950.
3. Gough, H. G. Identifying psychological femininity. *Educ. psychol. Measmt*, 1952, **12**, 427-439.
4. Gough, H. G. What determines the academic achievement of high school students? *J. educ. Res.*, 1953, **46**, 321–331.
5. Hathaway, S. R., & McKinley, J. C. *The Minnesota Multiphasic Personality Inventory.* Minneapolis: Univer. of Minnesota Press, 1943.
6. Kelly, E. L., & Fiske, D. W. *The prediction of performance in clinical psychology.* Ann Arbor: Univer. of Michigan Press, 1951.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

A Repetition of Meyer Williams' Study of Intellectual Control under Stress and Associated Rorschach Factors[1]

Virgil R. Carlson and Richard S. Lazarus
The Johns Hopkins University

For many years psychologists interested in problems of psychological stress have assumed that impairment of performance as a consequence of stress was associated entirely with such personality variables as neuroticism and lack of emotional control. This point of view was supported in a study by Meyer Williams [12], which was aimed at the validation of two indices of "emotional" or "intellectual" control in the Rorschach test. Combining these two indices in the prediction of performance on the Digit-Symbol test under conditions of psychological stress, Williams obtained a remarkably high multiple correlation of .824. This study was important not only because it provided support for certain Rorschach concepts, but also because it related directly to the practical and theoretical problem of predicting behavior under conditions of psychological stress.

There are four principal reasons why this finding of Williams is surprising and requires verification. Some of these considerations have been pointed out in a recent review article on the field of stress [8]. In the first place, it is impossible to define psychological stress in terms of stimulus situations such as threat of failure, electric shock, and similar factors, when the performance of particular individuals is being predicted. Varying motivations and past experience of the subjects must determine the extent to which they are involved in the situation, how they interpret it, and what their response will be. Second, Williams' stress situation was a combination of stresses, including simultaneously the threat of failure, electric shock, and the presence of observers watching the subjects as they performed. The situation therefore tapped a variety of needs and past experiences which are not likely to be comparable from subject to subject. Any single personality measure such as "emotional control" cannot be expected to correlate highly with responses to so complex a threat situation. Third, certain experiments have recently produced evidence which conflicts with the findings of Williams. For one thing, some of these experiments have shown that "neurotic" or "anxious" subjects may perform better than nondisturbed subjects in certain kinds of stress situations [5,10]. Also, a number of investigators have been unable to predict performance under stress by means of Rorschach measures even when the task used is comparable or the same [1, 4, 6, 9]. Finally, a serious limitation of Williams' study is the fact that it depended upon a sample of 25 cases, which is really too small for a reliable estimate of a correlation.

For these reasons it was important to determine whether Williams' results are repeatable. Since it is possible that his experiment may have contained just the right combination of factors for predicting performance under stress from Rorschach measures, we have made the greatest effort to repeat the study as accurately as possible in every essential detail.

Procedure

Twenty-five white male college students reported individually for two experimental sessions. In the first session the subjects were given the Rorschach test according to the procedure outlined in Beck [2]. After a five-minute rest period, five 90-second trials were given on the Wechsler-Bellevue Digit-Symbol test [11][2] with a one-minute break between trials. Each subject was told that the Digit-Symbol test is part of a standard test of intelligence, and he was encouraged to do as well as possible.

---

[1] This research was supported in part by the United States Air Force under Contract No. AF 33(038)-22624 monitored by the Perceptual and Motor Skills Research Laboratory of the Air Training Command Human Resources Research Center. Permission is granted for reproduction, translation, publication, use and disposal in whole and in part by or for the United States Government.

The second session occurred a day or two later. The subject was first given six more practice trials on the Digit-Symbol test. Again one-minute rest periods intervened between successive trials, and a three-minute rest period was interpolated between the third and fourth trials. The last three of these practice trials were designated the "control period" and constituted an estimate of the subject's nonstress performance. Then the subject was taken immediately to another room for three more Digit-Symbol trials under the stress conditions.

During the "stress period" the subject sat at a desk in a small room behind a one-way screen. He could not see out but could be observed from outside the room. A camera was mounted in front of the screen facing into the room. Three to five persons, including at least one woman, observed the subject's performance.[3] The subject was introduced to the observers beforehand and shown the one-way screen arrangement. Electrodes were strapped to the subject's nonwriting hand and a panel containing a white light and a red light was located on the table in front of him. The experimenters rated each subject, during his performance, on a four-point scale for behavioral signs of stress, such as trembling, sweating, tenseness, etc.

The following instructions were played on a tape recording:[4]

"You are now being observed by a number of psychologists on the other side of the screen who are taking notes and continuous photographs of all your reactions throughout the remainder of this experiment. All directions are to be followed implicitly. Rest your arm attached to the electrodes on the table and keep it there from now on. You will notice that the electrodes on your arm are connected to the panel before you. (White light turned on and left on.) The white light which has just gone on indicates that our shock apparatus has been turned on. You are connected to this apparatus. During the following period you may receive a strong electric shock whenever the observers feel that your test performance is not up to our standards. (Red light turned on.) Whenever the red light goes on, you are not meeting our standards and you are in danger of being shocked—like this! (Shock administered through electrodes, red light turned off.) Now pick up your pencil and look at the test blank before you. It is the same test you took just a few minutes ago. Your instructions are the same as before. At the signal 'Go' work as fast as you can until you are told to stop. Concentrate on your work. Remember, your reactions are being observed and continually photographed and you will be shocked whenever your test performance falls below our standards. Get set for Trial 1. Go!"

Five seconds after the "Go" signal the red light was turned on and left on until five seconds before the end of the trial. Just before the red light went off, the subject received a rather strong electric shock. Ninety seconds after the "Go" signal, the tape recording resumed:

"Stop! Place the test you have just completed on the right side of the table. You will shortly be called upon to repeat the test. Remember, you are being observed and continually photographed and you will be shocked whenever your test performance is below our standards. (Pause) Get set for Trial 2. Go!" This second "Go" signal occurred one minute after the "Stop" signal at the end of Trial 1. The same procedure was repeated for Trials 2 and 3. At the end of Trial 3, the tape recording informed the subject, "The experiment is finished. Remain seated until the electrodes are removed."

The subject was then taken to another room and asked to fill out Winne's Neuroticism Scale [13] and a questionnaire concerning reactions to stressful situations. After completing the questionnaires, he was interviewed in order to determine his subjective reactions to the experimental situation and to explain to him the purpose of the experiment. He was also cautioned to maintain secrecy regarding the experiment. The questionnaire and interview sessions were additions to Williams' procedure.

A control group of 15 subjects was given the same number of Digit-Symbol trials according to the same time schedule. The stress conditions, of course, were not imposed on these subjects.

---

[2] The test was lengthened sufficiently to preclude the possibility of anyone completing it in the time allowed.

[3] Williams' group of observers consisted of three to five persons, including at least two women. Our one-way screen was approximately 2×3 feet, whereas Williams' screen was 8×4 feet. Our experimental room was also apparently smaller than the one used by Williams.

[4] These instructions were reproduced verbatim from Williams' report.

Table 1
Summary of Rorschach Performance

Rorschach category	Experimental group			Williams' experimental group		
	Mean	Range	*SD*	Mean	Range	*SD*
Sum C/Total C	.83	.5–1.0	.176	.88	.5–1.2	.20
$F+\%$ Total	76.8	50–100	13.31	81.6	70–100	6.59
$F+\%$ Color Cards	70.1	0–100	26.40	75.5	50–100	9.28

Results

Each of the Rorschach records was scored both according to Beck [2] and according to Klopfer and Kelley [7], yielding a basis for four quantitative scores. These scores were: (*a*) Sum C/Total C, (*b*) $F+\%$ for the total record, (*c*) $F+\%$ for the color cards alone, and (*d*) Buhler, Buhler, and LeFever's Basic Rorschach Score [3].[5] The first three of these were the Rorschach measures used by Williams.

Sum C/Total C was determined by summing the weighted numerical values of the various color responses ($FC = .5$, $CF = 1.0$, $C = 1.5$) and dividing the sum by the total number of color responses. A high Sum C/Total C value represents poor integration of form with color, and supposedly, poor emotional control. $F+\%$ is the ratio (multiplied by 100) of the number of responses scored $F+$ to the number scored $F+$ plus the number scored $F-$. A high $F+\%$ represents good form quality, and, supposedly, a high degree of intellectual control. The Basic Rorschach Score is derived from a diagnostic sign list based primarily on Klopfer's system of scoring. Some 102 separate signs are each given a numerical weight, positive or negative. The algebraic sum of these weights constitutes the Basic Rorschach Score, representing an over-all estimate of the degree of maladjustment or personality disintegration. A high positive Basic Rorschach Score is said to indicate good personality integration.

Table 1 compares the performance of our experimental group with the performance of Williams' group on the pertinent Rorschach variables. The important thing to note here is that the two groups appear to be quite comparable. If anything, our group represents a somewhat greater range of individual differences, at least with respect to the $F+\%$.

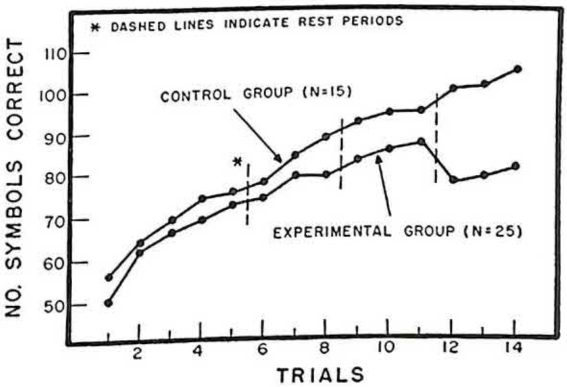


Fig. 1: Digit-symbol test performance for the control and experimental groups.

Comparison of the two groups in performance on the Digit-Symbol test is somewhat more complicated. Williams found that 11 trials were sufficient to reach a plateau of no further improvement. The mean number of digits correctly completed for Trials 9, 10, and 11 (Control Level) minus the number correctly completed for Trials 12, 13, and 14 (Stress Level) was taken as a measure of decrement under stress. Figure 1 shows that our control group certainly, and our experimental group probably, had not reached a plateau by Trial 11. Consequently, Williams' stress score more nearly represents a decrement from the point of maximum performance than does ours. Most likely this accounts for the fact that Williams' average Stress Decrement is nearly twice as large as ours (refer to Table 2). If we were to take the extrapolated learn-

---

[5] In order to conform with Williams' procedure our Rorschachs were administered according to Beck's instructions, which differ from the procedure set forth by Buhler, Buhler, and LeFever. Our obtained Basic Rorschach Scores may therefore lack strict comparability with those obtained by other investigators.

Table 2
Summary of Digit-Symbol Test Performance

Digit-Symbol measures	Control group		Experimental group	
	Mean	*SD*	Mean	*SD*
(1) Control Level	93.7	18.97	85.1	16.69
(2) Stress Level *	102.1	19.70	79.3	15.19
Stress Decrement (1 minus 2)	−8.4†	4.04	5.8‡ (10.4) §	8.34 (5.70)
t .	7.72		3.41	
p	.001		.01	

* A second "control level" for the control group.
† Minus score represents improvement in performance.
‡ If the extrapolated learning curve is used instead of the actual control level, the decrement is 11.0.
§ Values in parentheses are those obtained by Williams.

ing curve of our experimental group as an estimate of the group's maximum performance, we would obtain an average Stress Decrement which is almost exactly equal to the value obtained by Williams. Our actual stress measure, however, is statistically significant and its standard deviation is fully as large as in Williams' data. Further, since our control group shows a significant increase in performance during the period corresponding to the stress trials of the experimental group (see Table 2), we believe that we succeeded in producing as effective a stress situation as did Williams.

In our data the amount of Stress Decrement on the Digit-Symbol test is correlated with the Control Level. This was not the case in Williams' data, and probably the reason is that his subjects had more nearly approached their maximum level of performance. In any event, we derived a set of "corrected" Stress Decrement scores which do not correlate with performance in the control period. We first calculated the predicted Stress Decrement scores by means of the regression equation predicting Stress Decrement from Control Level. Then the obtained Stress Decrement minus the predicted decrement constituted the Corrected Stress Decrement scores.[6] Correlations with the various Rorschach and other personality measures were computed for this measure as well as for the original Stress Decrement. The same result, elimination of the influence of control level of performance, could have been achieved by computing partial correlation coefficients, and any correlation involving Corrected Stress Decrement should be interpreted as a partial correlation.

We found, as did Williams, that subtracting the lowest trial in the stress period from the highest trial in the control period resulted in a set of scores (here called Stress Maximum) which correlated highly with the Stress Decrement (differences between the means of the two periods). Consequently, there was little point in correlating this measure with other variables. Another measure (not used by Williams), however, correlated only slightly with Control Level or with Stress Decrement and so seemed worth while to explore. This is im-

Table 3
Intercorrelations among Digit-Symbol Test Measures

	Stress Decrement	Corrected Stress Decrement	Improvement * under Stress	Stress † Maximum	Stress Level
Control Level	.42 (.05)‡	−.01	.09	.51	.87
Stress Decrement		.90	.33	.94 (.93)	
Corrected Stress Decrement			.32		

* Last stress trial minus first stress trial.
† Highest control trial minus lowest stress trial.
‡ Values in parentheses are those obtained by Williams.

---

[6] Another method, which is perhaps more logical, consists in predicting stress-level performance from the control level by means of the regression equation predicting Stress Level from Control Level and then subtracting the Stress Level scores from the predicted scores. The scores obtained in this way correlated .996, however, with the Corrected Stress Decrement scores described above, suggesting that it made little difference which method was used.

provement under stress. It was obtained by subtracting the score for the twelfth trial (first trial under stress) from the score for the fourteenth trial (last trial under stress).

The intercorrelations of these various Digit-Symbol measures are presented in Table 3. A correlation of about .40 is required for significance at the 5% level of confidence.[7]

The correlations between the stress measures and the Rorschach measures are shown in Table 4. The two most important correlations in Williams' study were between Stress Decrement on the one hand and Sum *C*/ Total *C* and *F* + % on the other. The combined predictive power of these two Rorschach measures was relatively great, yielding a multiple correlation of .824. We do not corroborate this finding at all. Our correlations between Stress Decrement and *F* + %, and *F* + % on the color cards, are near zero. Our correlation between Stress Decrement and Sum *C*/Total *C* approaches significance, but it is in the opposite direction from Williams' correlation. Moreover, this last correlation becomes clearly insignificant when the influence of control performance is removed from the Stress Decrement score. As might be expected, our multiple correlation between Stress Decrement and Sum *C*/Total *C* plus *F* + % (.39) is practically the same as the correlation between Stress Decrement and Sum *C*/ Total *C* alone. It might be considered to approach significance except for the fact that it is in the wrong direction. In addition, Stress Decrement failed to correlate with the Basic Rorschach Score, the Behavior Rating, or either of the questionnaires.

Table 4

Correlations between Digit-Symbol Test Measures and Rorschach Measures

	Sum *C* / Total *C*	*F*+% Total	*F*+% Color cards	Basic* Rorschach score
Stress Decrement	–.37 (.35)†	.16 (–.61)	.06 (–.72)	–.17
Corrected Stress Decrement	–.28	.05	.02	–.07
Improvement under Stress	–.07	.12	.03	–.18

* Buhler, Buhler, and LeFever sign check list.
† Values in parentheses are those obtained by Williams.

The only positive instance of correlation in all our data was between Improvement under Stress and Winne's Neuroticism Scale. The correlation coefficient in this case was + .54, meaning that the subjects showing greater neuroticism or anxiety on the inventory tended to show more improvement during the stress trials. Also, the correlation between the neuroticism inventory and Behavior Rating[8] approached significance with a coefficient of + .37. Neither the inventory nor the Behavior Rating correlated with any of the Rorschach measures, however.

The most noteworthy fact derived from interviewing the subjects was that they considered the electric shock the most disturbing factor in the situation. Most of them expressed the view that the one-way screen helped them to ignore the presence of the observers. And in general they claimed to have paid little or no attention to the red light, which was supposed to signify below-standard performance. When they did think about the light, it seemed to carry the implication of the imminence of shock rather than failure. It would take some further experimentation to ascertain the actual contribution of each of these factors to the decrement produced in performance. In our own opinion the electric shock did indeed contribute the most, but it is not clear in what precise way. Part of the effect of the shock was apparently just a simple and momentary mechanical interruption in performance.

Discussion

The primary result of the present study is the complete lack of corroboration of Williams' findings. This result, however, agrees with our theoretical expectations concerning the problem of psychological stress and with other experiments having to do with the prediction of performance under stress from Rorschach measures. Experiments by Deese and Lazarus [4], Eriksen, Lazarus, and Strange [6], Stopol [9], and Allee [1] have all produced negative results. In the experiment most similar to Williams' prior to the present report, Stopol found no significant relationship between any Rorschach measure and performance under stress. Since Williams' experiment stands alone, it seems reasonable to reject his results as representative of the relationship between Rorschach variables and performance under stress.

---

[7] All correlations referred to in this paper are Pearson product-moment coefficients.

[8] Each of the experimenters (independently) rated the behavior of the subjects for overt signs of stress during performance on the stress trials. The rating categories ranged from 1 (little indication of stress effect) to 4 (pronounced indications of stress). Since the two independent sets of ratings agreed fairly highly, they were combined into a single Behavior Rating score.

Williams' experiment was undertaken from the point of view of validating certain Rorschach concepts, whereas in the researches cited above the Rorschach measures were secondary to the study of the effects of stress on performance and their prediction. From the viewpoint of Rorschach validity, even if form quality and form-color integration could be properly designated as representing emotional or intellectual control, our theoretical position with respect to the determinants of performance under stress would not lead us to expect a substantial relationship between such "control" and performance under stress. Indeed, one could even make the reverse prediction on the basis of the findings that neurotic individuals may do better than nonneurotics under certain kinds of stress conditions [5, 10].

What is important from an empirical point of view is that these Rorschach measures do not enable us to predict the kinds of behavior which have been studied here. We cannot be certain whether this means that the Rorschach concepts are inadequate or whether the behavior studied is inappropriate to test them. In order to be useful, the concepts of emotional and intellectual control must lead to some valid behavioral prediction.

A final word must be offered with respect to the correlation which was found between Improvement under Stress and the scores on the neuroticism inventory. Although this correlation is large enough to be statistically significant, in a matrix of so many correlations one might obtain such a relationship by chance. But if it can be substantiated, it is undoubtedly related to the finding, mentioned previously, that neurotics sometimes do better than nonneurotics in stress situations.

Summary

An experiment by Meyer Williams, in which a high degree of relationship was found between performance under stress and Rorschach measures of emotional and intellectual control, was repeated with careful attention to every essential detail. The repetition was necessitated by the appearance of conflicting evidence concerning the efficacy of the Rorschach in predicting performance under stress, the belief that a competent theory of psychological stress precluded the finding of substantial relationships, and the fact that Williams' conclusions depended upon a small sample of 25 cases.

The task used was the Wechsler-Bellevue Digit-Symbol test. Each of 25 subjects completed a total of eleven trials on this test under control conditions and then three trials under stress conditions. The stress consisted of electric shock, failure information, and the presence of several people observing the performance of the subject. The difference between the subject's performance during the three trials preceding the stress period and the three stress trials constituted the measure of decrement due to stress. These data were compared with the performance of a control group, indicating stress effects which were quite comparable with those reported in Williams' study.

The performance of our group of subjects on the relevant Rorschach variables was also comparable to that of Williams' subjects with respect to mean, range, and standard deviation. There was no relationship, however, between the Rorschach measures and decrement in performance under stress.

In view of the fact that all other similar investigations have produced negative results, it is reasonable to conclude that the correlation reported by Williams is not representative of the relationship between Rorschach variables and performance under stress.

Received November 17, 1952.

References

1. Allee, Ruth. Rorschach responses of extreme deviation in an experimental stress condition. Unpublished master's thesis, Univer. of Missouri, 1948. Cited in F. McKinney, G. B. Strother, Ruth R. Hines, and Ruth A. Allee. Experimental frustration in a group test situation. *J. abnorm. soc. Psychol.*, 1951, **46**, 316-323.
2. Beck, S. J. *Rorschach's test.* I: *Basic processes.* New York: Grune & Stratton, 1944.

3. Buhler, Charlotte, Buhler, K., & LeFever W. D. Development of the basic Rorschach score with manual of directions. Los Angeles: Rorschach Standardization Studies, No. 1, 1948.
4. Deese, J., & Lazarus, R. S. The effects of psychological stress upon perceptual-motor performance. Human Resources Research Center Bulletin, in press.
5. Deese, J., Lazarus, R. S., & Keenan, J. Anxiety, anxiety-reduction, and stress in learning. Human Resources Research Center Bulletin, in press.
6. Eriksen, C. W., Lazarus, R. S., & Strange, J. R. Psychological stress and its personality correlates. II: The Rorschach test and other personality measures. *J. Pers.*, 1952, **20**, 277-286.
7. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1942.
8. Lazarus, R. S., Deese, J., & Osler, Sonia. The effects of psychological stress upon performance. *Psychol, Bull.*, 1952, **49**, 293-317.
9. Stopol, M. S. An experimental investigation of the consistency of stress tolerance and related Rorschach factors. Unpublished doctor's dissertation, Columbia Univer., 1951.
10. Taylor, Janet A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81-93.
11. Wechsler, D. *The measurement of adult intelligence.* (3rd. Ed.) Baltimore: Williams & Wilkins, 1944.
12. Williams, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1947, **11**, 21-29.
13. Winne, J. F. A scale of neuroticism: An adaptation of the Minnesota Multiphasic Personality Inventory. *J. clin. Psychol.*, 1951, **7**, 117-122.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Subcategories of Rorschach Human Movement Responses: A Classification System and Some Experimental Results[1]

Robert P. Barrell
Veterans Administration Hospital, New Orleans, Louisiana

Several authors, in Rorschach literature, point to some relationship between various types of human movement responses (M) and various personality characteristics, and present clinical hypotheses regarding the interpretation of them. Rorschach [10] himself had some ideas regarding the relationship of extensor and flexor aspects of the movement to activity and passivity within the personality of the subject. Wheeler [13] seems to feel that certain characteristic types of M are found in overt homosexuals. Beck [3] suggests that the direction of the movement (away from or towards the center of the card) tells something about the subject's activity or passivity. Piotrowski [9] differentiates several types of movement (along an extensor-flexor scale) as being related to various personality characteristics, relative to self-assertion and submissiveness.

It is common practice in the clinic to utilize the hypotheses referred to above, or similar ones, regarding the meaning of the various types of human movement responses, and to accept them as maxims for interpretation. Yet, aside from the work of Wheeler, there does not seem to be any clear validation of these hypotheses. Thus the problem demands investigation.

The Classification System

It is obvious that for crucial experimental work there must be a clear-cut classification system. This system, it seems, should meet at least two criteria: (*a*) The various categories should be clearly defined, permitting the least possible amount of ambiguity; (*b*) the subcategories should yield a high coefficient of reliability with respect to scoring. Another desirable aspect of such a classification system would be that there be, insofar as possible, a continuum instead of purely categorization, which would allow for more adequate statistical manipulation.

Yet, when the existing classification systems for M are examined, not even one is found clearly to meet these criteria. Furthermore, it is obvious that these classification systems overlap each other to a great extent. The Hutt-Milton classification [8] is similar to that of Rorschach [10] in that extensor and flexor aspects are considered; Werner [12] draws closer in his "action," "expression," and "posture" classification; and Zubin [14, 15] cuts across most of these in a different way, considering not only activity but interaction with other people and with the environment in general.

Thus, for purposes of testing some hypotheses regarding the subcategories of M [1] it was decided to develop a classification system that would attempt to systematize all of the elements of the existing breakdowns as well as meet the criteria set forth above.

Following Zubin [14, 15] the new classification system of the M response was built around two concepts prominent in the previous classifications: (*a*) the amount of *energy* expended in the movement; and (*b*) the amount of *interaction* with the environment shown by the response. Thus was developed, first, the *Energy Classification*, which involved a characterization of the response with respect to the amount of energy expended in terms of the total body, the limbs and the face, respectively. Next was developed an *Interaction Classification*, which considered: (*a*) degree of interaction with nonanimate objects in the environment; (*b*) degree of interaction with humans specified or implied by the response; and (*c*) the type of interaction with humans specified or implied by the response. Where there was no interaction with humans or with nonhuman objects in the environment, a classification of the *kind* of action involved in the response was used; for example, "a man standing" would be classified as involving no movement in space, and "an angel flying" would show no specific direction of movement.

---

[1] From doctoral dissertation, University of Michigan, 1950 [1]. The author wishes to acknowledge his indebtedness to the late Professor Louis Granich and to Professors Max L. Hutt, E. Lowell Kelly, Irving H. Anderson, Edward S. Bordin, and Donald G. Marquis.

Further, since it was desired to determine whether sex designation in the *M* response had any relationship to personality variables, a *Sex Designation Classification* was constructed. This consisted of: (*a*) designation as male and how so designated (whether primary or secondary sex characteristics, or both, were referred to, or whether no reason was given for the sex designation); (*b*) designation as female and how so designated; (*c*) absence of sex designation; (*d*) vacillation in sex designation, such as giving one sex in the free association and another in the inquiry, or verbal evidence of confusion concerning the sex of the person seen.

Details of the Classification System

The Energy Classification

1. *Whole body.* (*a*) Human seen as expending energy vigorously with the total body. Examples: fighting, dancing, running, whirling. (*b*) Human seen as expending a modicum of energy. Examples: walking, standing, talking. (*c*) Human seen as expending little energy. Examples: sitting, lying, kneeling; any postural response.

2. *Limbs.* (*a*) *H* or *Hd*,[2] seen as expending energy vigorously with limbs. Examples: hitting, kicking, throwing. (*b*) *H* or *Hd*, seen as expending modicum of energy with the limbs. Examples: shaking hands, playing patty-cake, stirring. (*c*) *H* or *Hd*, seen as expending little energy with the limbs. Examples: pointing, gesturing.

3. *Degree of activity in facial expressions* (regardless of *H* or *Hd*). (*a*) Much energy utilized. Examples: calling for help, shouting. (*b*) Modicum of energy utilized. Examples: laughing, talking, grimacing. (*c*) Little energy utilized. Examples: looking, smoking, peering, looking angry.

The Interaction Classification

1. *Objects in the environment.* (*a*) Human seen as not responding to any object in his environment. Action is entirely self-referred. (*b*) Human attempting to manipulate object or objects in his environment. Examples: pulling, pushing, lifting, carrying. (*c*) Human seen as avoiding object in his environment. Examples: running away from something, cringing. (*d*) Human seen as being manipulated by his environment. Examples: pushed, driven, etc. by some object in his environment (object specified). (*e*) Human just responding to some object in his environment, as holding something. (*f*) Human looking at something in his environment.

2. *Degree of interaction with people.* (*a*) Human not seen as relating to other people in his environment. Action is entirely self-referred. (*b*) Human aware of the presence of other people in his environment, but not taking any direct action as a result of this. Examples: looking at someone, smiling at someone, shielding something. (*c*) Human seen as responding to another human, but amount of action is only moderate. Examples: walking, talking, begging. (*d*) Human involved in much action with other people. Examples: fighting, dancing, working.

3. *Type of interaction with people.* (*a*) Expression of aggression. (*b*) Expression of affection. (*c*) Cooperation with others, as in work activity. (*d*) Subject of aggression. (*e*) Miscellaneous activity, e.g., facing each other. (*f*) Lecturing, advising, talking, gossiping, directing. (*g*) Vacillation, e.g., affection-aggression.

4. *Direction of movement.* (*a*) Movement without any specific direction. Examples: floating in air, keeping time with music. (*b*) Moving up. (*c*) Moving down. (*d*) Moving to the right. (*e*) Moving to the left. (*f*) No movement in space, e.g., standing, lying.

Sex Designation Classification

1. *Human seen as male.* (*a*) No reason given. (*b*) Secondary sex characteristics given, e.g., clothing, shape. (*c*) Primary sex characteristics given. (*d*) Both primary and secondary sex characteristics given.

2. *Human seen as female.* (*a*) No reason given. (*b*) Secondary sex characteristics given, e.g., clothing, shape. (*c*) Primary sex characteristics given. (*d*) Both primary and secondary sex characteristics given.

3. *Vacillation.* Sees male, then female, or vice versa.

4. *No sex designation.* No indication of the sex of the figure seen is presented in the response.

Scoring Procedure

Every *M* response in the subject's Rorschach record is scored for the presence or absence of

---

[2] *H* is the conventional symbol for *Human*, and *Hd* for *Human Detail*.

each of the subcategories defined and grouped above. Thus the scores on these subcategories used for statistical purposes are the sum of the subject's scores on each subcategory for all of his *M* responses. For example, Subject X's score on the subcategory *Body a* (much expenditure of energy with the total body) would be the sum of all of X's *M* responses in which the person was seen as expending much energy with the total body. Similarly, his score on subcategory *Type a* (expression of aggression) is the sum of all his *M* responses in which the expression of aggression appears. Obviously, the subcategories can be combined into groups and scores can be obtained for the groups, as, for example, the total number of Male designations. Subject X's score here would be, of course, the total number of *M* responses in which he designated the sex as male.

Reliability of Scoring

In order to obtain some measure of the reliability of the scoring, a sample of the records from the study [1] was given to five clinical psychologists whose Rorschach training and experience roughly paralleled that of the author. Prior to the "test scoring" 2 trial records were chosen, 1 being a long one, 1 a short one. Each scorer was trained by the author in the scoring of the above subcategories of *M* on these 2 records, and the results of their scoring here were not included in the reliability study. Subsequently each individual scored 3 different Rorschach records, selected in random fashion. Thus there were available two sets of scores for each of the 15 Rorschach records: one obtained by the author and the other by one of the scorers referred to above. Rank-order coefficients were computed between the author's scoring and the scoring of the other psychologists for each of the *M* subcategories. Table 1 presents the results.

Table 1

Rho Coefficients for Interscorer Reliability

M Subcategory	Rho	*M* Subcategory	Rho
Body *a*	.849	Type *b*	.799
Body *b*	.700	Type *c*	.618
Body *c*	.621	*Type *d*	......
Total Body	.889	Type *e*	.555
*Limbs *a*	......	Type *f*	.691
Limbs *b*	.363	*Type *g*	......
Limbs *c*	.610	Direction *a*	.612
Total Limbs	.524	*Direction *b*	......
Face *a*	.705	*Direction *c*	......
Face *b*	.675	*Direction *d*	......
Face *c*	.597	*Direction *e*	......
Total Face	.627	Direction *f*	.757
Objects *a*	.915	Male *a*	.821
Objects *b*	.453	Male *b*	.923
*Objects *c*	......	*Male *c*	......
*Objects *d*	......	*Male *d*	......
Objects *e*	−.083	Total Male	.863
*Objects *f*	......	Female *a*	.741
People *a*	.875	Female *b*	.403
People *b*	.665	*Female *c*	......
People *c*	.572	*Female *d*	......
People *d*	.284	Total Female	.855
Type *a*	.524	Sex–Vacillation	1.000
		Sex–No designation	.782

* Omitted from study [1] because of insufficient frequency of occurrence in sample to permit statistical manipulation. Hence, omitted here.

When the rho coefficients of Table 1 are converted into r's and the significance is computed, 5 of them fall below the 5% level of significance (rho of .480) and 10 below the 1% level of significance (rho of .60). However, 23 of the 33 relationships yield signficances beyond the 1% level, thus indicating that moderate agreement can be obtained among scorers who have had only a minimum amount of training.

The Experimental Results

The following two generalized hypotheses were constructed for purposes of investigating the relationships between the subcategories of *M* and personality characteristics: (*a*) The various subcategories into which the *M* responses may be subdivided are differentially related to various categories of intellectual functioning; (*b*) various categories of *M* responses show significant relationships to different categories of emotional behavior.

Subjects and Experimental Data

The subjects for our study of the above hypotheses were 121 males who were beginning graduate students in clinical psychology, having graduated from a recognized college or university, and specially selected, from a large group of applicants, for training in the Veterans Administration Training Program for Clinical Psychologists. These subjects had been studied intensively for a period of one week at

Ann Arbor, Michigan, during the summer of 1947, as part of the Veterans Administration Assessment Program for Clinical Psychologists.[3] During this period they had been given many objective, projective, and situational tests, and they were subjected to intensive interviews. There were also available records of scholastic achievement and letters of recommendation. Upon the basis of all of these data, each subject was rated by a staff of experienced clinical psychologists and psychiatrists on a number of intellectual and emotional traits. These ratings are taken as criteria in our study, since we consider them to be reasonably valid measures of what the subjects were like at the time of assessment.

Procedure

1. The 121 Rorschach records were scored for *M* according to procedures outlined by Beck [2] and for the subcategories of *M* according to our new classification system.

2. The frequency distribution for each *M* subcategory was prepared, and the means and standard deviations were computed for each. Fourteen of the 47 subcategories had to be omitted from the study at this point because of insufficient frequency of occurrence in our sample (less than 5 cases per cell) for use of the chi-square technique.

3. Frequency distributions were made, and means and standard deviations were computed for the ratings of the 121 subjects on the various intellectual and emotional traits.

4. The two generalized hypotheses above were restated in terms of the individual subcategory classifications and the individual rating-scale items, the subhypotheses thus constructed being designed to test the generalized hypotheses.

5. The Rorschach variables were divided into two or three categories: (*a*) high and low, with the division as near the mean as possible (approximately 50% in each group); (*b*) high, medium, and low (approximately 33-1/3% in each group). These divisions were made according to the distribution of the scores for each variable. The ratings on the various traits utilized as our criterion variables were similarly divided, but only into high and low categories, with their division as near the mean as possible, dependent upon the shape of the distribution.

6. Chi squares were then computed for the various *M* subcategories and the various rating-scale items, the chi squares computed being chosen to test the hypotheses presented above.

7. Scattergrams were plotted for all of the relationships computed in Step 6 that had a p value of less than .20, and the shape of the curve was noted.

8. Pearson r's, or etas, dependent upon the shape of the curve in the scattergram, were computed for all the relationships shown by the chi square to have a p value of less than .20.

Table 2

Level of Significance of the Correlations with Intellectual Variables

M Sub-Category	Broad Interests	Imaginative	Independent Minded	Quality of Intellectual Accomplishments
Body *a*	....	....	.01	.01*
Body *b*	.05	.05	....	.05
Body *c*	....	.01*	.01	.01*
Total Body	.01*	.01*	.01	.01*
Limbs *c*	....	....	....	....
Total Limbs	....	....	....	....
Face *a*	....	....	....	....
Face *b*	....	....	....	.05
Face *c*	....	....	.01	.05
Total Face	....	.05	.01	.01
Object *a*	.01*	.01*†	.01	.01*
People *a*	.01	.01*	.01	.01*
People *b*	....	....	.05	.01*
People *c*	....	.01*†	....	.01
Type *a*	....	....	....	....
Type *b*	....	....	....	....
Type *c*	.01*	....	.05	.01*
Type *e*	....	....	....	....
Type *f*	....	....	....	.05
Direction *a*	....	....	....	....
Direction *f*	.01	.01	.01	.01
Male *a*	.01	.01	.01	.01
Male *b*	....	.05	....	.05
Total Male	....	.01*	.01	.01*
Female *a*	....	....	....	....
Total Female	....	....	....	.01
Sex–Vacillation	....	....	....	....
Sex–No designation	.01	.01*	.05	.01*

* Correlation with **R** partialed out is significant at the 5% level or better.

† Correlation with **M** partialed out is significant at the 5% level or better.

---

[3] This program is described in detail by Fiske [4], Kelly [5], and by Kelly and Fiske [6, 7].

This p value was chosen as an arbitrary limit for further exploration since it was believed that when the chi square p values were higher than .20, the probability that a significant correlation existed was very small.

9. To rule out the effect of the total number of responses, R, and the total number of M responses on the results, partial r's, partialing out the effect of R and M, respectively, were computed for all relationships shown by Step 8 to have a correlation significant at the 5% level or greater.

Results

1. *Relationships Between* M *Subcategories and Intellectual Functioning*

Table 2 presents the results relative to the relationships between the M subcategories and intellectual functioning. Here the hypothesis cannot be rejected. Fifty-two of the 112 (zero order) correlations of the subcategories with the intellectual variables are significant at the 5% level or better. Eight of the 28 subcategories are significantly related to "Broad Interests"; 12 of the 28 are significantly related to "Imaginative"; 13 are significantly related to "Independent Minded"; and 19 of the 28 are significantly related to "Quality of Intellectual Accomplishments." Nine of the subcategories are not significantly correlated with any of the intellectual variables.[4]

The subcategories of M, then, are differentially related to the intellectual variables. They tend to correlate with these variables independently of R, but are not independent of R with respect to "Independent Minded." Nineteen of the 52 relationships significant from zero-order correlation remain significant when the effect of R is held constant, but none of these involves a relationship with "Independent Minded." These partial correlation coefficients range from .175 to .601, accounting for from 3 to 36 % of the variance.

It is questionable whether M should be partialed out, as is R, since a score on M inevitably means a score on a subcategory. However, when M is partialed out, we find that the subcategories are independent of M in their predictions in only two cases, which could have occurred by chance.

Since less than 50% of the subcategories of M are independent of R in their predictions of the intellectual variables and only two are independent of M (which could have occurred by chance) in these predictions, it seems that the subcategories are not promising variables for the clinician to rely upon in assessing intellectual abilities.

2. *Relationships between* M *Subcategories and Emotional Variables*

The hypothesis that the various subcategories are differentially related to various emotional variables is not substantiated. Only nine of 132 correlations are significant at the 5% level or better, and only 1 significant at the 1% level or better. This is hardly above chance expectancy, since it would be expected that approximately 7 out of 132 correlations would be significant at the 5% level or better by chance alone. Only 4 of the relationships are independent of R, and 2 independent of M. These correlations could have been obtained by chance.

It may be argued that these low correlations were obtained because of low reliabilities in the criterion traits. However, Kelly and Fiske [7] and Tupes [11] present data on the reliabilities of the ratings. Estimated indexes of reliability of the individual rating-scale items utilized to test the above hypothesis are as follows: .80, .89, .81, .85, .84, .79, .79, .90, .85, .84, .88. Thus it does not appear that the low correlations obtained can be explained on the grounds of low reliability of the criterion data.

It may be that the possible selection of the population on the personality variables does not offer an optimal situation for an adequate test of the hypothesis. Nevertheless, it seems clear that one is not justified, with the type of subjects utilized in this study and with the type of classification of M responses employed, in concluding that the subcategories of M are related to the emotional variables investigated here.[5]

---

[4] These are: *Limbs c, Total Limbs, Face a, Type a, Type b, Type c, Direction a, Female a,* and *Vacillation in Sex Designation.*

[5] The variables are: "Readiness to Cooperate vs. Obstructiveness"; "Assertive vs. Submissive"; "Suspicious vs. Trustful"; "Generous vs. Self-centered"; "Cautious vs. Adventurous"; "Socially Poised vs. Socially Clumsy"; "Frank vs. Secretive"; "Social Adjustment"; "Appropriateness of Emotional Expression"; "Characteristic Intensity of Inner Emotional Tension"; and "Sexual Adjustment."

Implications

1. M *Subcategories and the Intellectual Variables*

Study of Table 2 shows some interesting relationships and provides for some speculations, to be investigated experimentally at a later date.

First, the *Body* subcategories correlate significantly with the intellectual variables in 12 out of 16 cases. Yet the *Limbs* subcategories do not correlate significantly with *any*, and the *Face* subcategories correlate significantly with only 6 of the 16. It is hypothesized that the *Limbs*, in particular, and also the *Face* subcategories represent something very personal to the subject, have very personal meaning for him. Hence, projection of *M* here is less related to intellect than projection of *M* into the total body, and more related to emotional factors. Along the line of this speculation, it is pertinent to note that in correlations between the subcategories and the emotional variables, the *Body* subcategories correlated significantly with 2 of the 18 variables tested whereas the *Limbs* subcategories correlated significantly in 3 of the 9 relationships tested.

Of greater interest, perhaps, is the following set of observations: the correlation of *Object a* and *People a* (where there is no interaction with objects or people in the environment) with the intellectual variables is significant in all cases, whereas the correlations of the other Interaction classifications are not significant in all cases; the *Cooperation* interaction correlates significantly with three of the intellectual variables, whereas only one other *Type* subcategory (*Lecturing*) correlates significantly with *any* of the intellectual variables and that with only one ("Quality of Intellectual Accomplishments"); the *No Designation of Sex* subcategory correlates significantly with *all* of the intellectual variables, whereas *Vacillation in Sex Designation* correlates with *none,* and designation as *Female* correlates with only one ("Quality of Intellectual Accomplishments"). The hypothesis growing out of these data is that the person who functions well intellectually is the one who isolates (in the Freudian sense) intellect from emotion; that is, emotional conflict, when not isolated, impairs intellectual functioning. The "isolator" would not be concerned with interaction with objects or people, and hence is not particularly concerned regarding the sex of the person seen on the Rorschach nor with highly emotionally charged interaction, projecting only "cooperation."

A third hypothesis growing out of the data is based on the fact that there is more significant correlation between the designation as *Male* and the intellectual variables than there is with the other sex-designation subcategories. The suggestion is that the man who gives much *Male* designation over the other sex-designation subcategories, does so because he has worked out a good masculine identification, and in many ways then can be said to be "better adjusted" than those males who have not worked out a good masculine identification. Hence, he is able to utilize his intellect more effectively.

2. M *Subcategories and the Emotional Variables*

Since the correlations are little above chance, it does not appear that it is possible, at the present time, to utilize the various subcategories of *M* employed here as measures of the emotional traits investigated. Perhaps these subcategories are related to some of the traits not investigated here. Perhaps, also, some combination of these Rorschach variables will be good predictors of these emotional traits. But, at the present stage of investigation, it seems to be a very dubious procedure for the clinician to utilize the subcategories utilized here as measures of the emotional variables tested here. In the event of lack of evidence concerning the relationship between other types of subcategories of *M* and other emotional traits, the negative results here emphasize that the clinician should be doubly cautious about making interpretations from *M* subcategories. It is true, of course, that the clinician may obtain clues to be investigated with other Rorschach data and with other clinical data, as he does with many other clinical hunches. But to assume that there is anything like a one-to-one relationship between the *M* subcategories and emotional variables, or to assume that the *M* subcategories have any clear validity as measures of emotional traits, seems to be an extremely hazardous procedure.

Summary

1. The necessity for experimental verification of hypotheses regarding the interpretation of various types of human movement responses in the Rorschach test is pointed out.

2. A new classification system for *M* responses, embodying elements of existing systems, is developed, in an effort to meet requirements of good experimental work, and presented here.

3. An experimental study of these subcategories of *M* is reported, investigating the following generalized hypotheses: (*a*) The various subcategories into which the *M* responses may be subdivided are differentially related to various categories of intellectual functioning; (*b*) various categories of *M* responses show significant relationships to different categories of emotional behavior.

4. Subjects utilized for this study were 121 males, beginning graduate students in clinical psychology, studied intensively in the Veterans Administration Assessment Program for Clinical Psychologists.

5. Criteria for the study were the ratings of the 121 subjects on a variety of intellectual and emotional traits, made by a staff of experienced clinical psychologists and psychiatrists.

6. Results: Hypothesis (*a*) was substantiated by the data, although it was discovered that the subcategories are not *independently* related to the intellectual variables; hypothesis (*b*) was *not* substantiated by the data.

7. Implications of the findings, particularly as concerned the clinician's use of *M* subcategories in interpretation of the Rorschach, are discussed.

Received December 22, 1952.

References

1. Barrell, R. P. The relationship of various types of movement responses in the Rorschach test to personality trait ratings. Unpublished doctor's dissertation, Univer. of Michigan, 1950.
2. Beck, S. J. *Rorschach's test.* Vol. 1. *Basic processes.* New York: Grune & Stratton, 1944.
3. Beck, S. J. *Rorschach's test.* Vol. 2. *A variety of personality pictures.* New York: Grune & Stratton, 1946.
4. Fiske, D. W. Consistency in the factorial structure in personality ratings from different sources. Unpublished doctor's dissertation, Univer. of Michigan, 1948.
5. Kelly, E. L. Research on the selection of clinical psychologists. *J. clin. Psychol.,* 1947, **3**, 39-42.
6. Kelly, E. L., & Fiske, D. W. The prediction of success in the V.A. training program in clinical psychology. *Amer. Psychologist,* 1950, **5**, 395-406.
7. Kelly, E. L., & Fiske, D. W. *The prediction of performance in clinical psychology.* Ann Arbor: Univer. of Michigan Press, 1951.
8. Milton, E. O. The influence of varied experimental sets upon certain Rorschach variables II: Stability of the human movement variable. Unpublished doctor's dissertation, Univer. of Michigan, 1949.
9. Piotrowski, Z. A. Tentative Rorschach formulae for educational and vocational guidance in adolescence. *Rorschach Res. Exch.,* 1943, **7**, 16-27.
10. Rorschach, H. *Psychodiagnostics.* Bern: Huber, 1942.
11. Tupes, E. C. An evaluation of personality-trait ratings obtained by unstructured assessment interviews. *Psychol. Monogr.,* 1950, **64**, No. 11 (Whole No. 317).
12. Werner, H. Perceptual behavior of brain injured mentally defective children: An experimental study by means of the Rorschach technique. *Genet. Psychol. Monogr.,* 1945, **31**, 51-110.
13. Wheeler, W. M. An analysis of Rorschach indices of male homosexuality. *Rorschach Res. Exch.,* 1949, **13**, 97-126.
14. Zubin, J., Chute, E., & Verniar, S. Psychometric scales for scoring Rorschach test responses. *Charact. & Pers.,* 1943, **11**, 277-301.
15. Zubin, J., & Young, Kathleen M. *Manual of projective and cognitive techniques.* Madison, Wisc.: College Typing Co., 1948.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Further Investigation of Rorschach Determinants Subjected to Factor Analysis

Harold L. Williams and James F. Lawrence
Letterman Army Hospital

Utilizing the Thurstone method of centroid analysis, Wittenborn [4, 5] has recently published two studies investigating the patterns of Rorschach scores obtained from two groups of subjects, college students and psychiatric patients. These factor analytic studies were undertaken to evaluate four basic hypotheses [4]: (*a*) the factorial composition for *M* will be different from the factorial composition for determinants involving the use of color; (*b*) the *W, CF, C, c,* and pure diffusion categories are so intercorrelated as to yield a factor differentiated from *F, M,* and *D*; (*c*) there is a productivity factor which (*d*) is involved to varying degrees in the several response categories. In general the hypotheses were confirmed, and his data indicated that *FC,* considered factorially, is more similar to *M* than to *CF* and *C*. His second study revealed some differences between the student and patient groups, introducing a new factor which the author believed was related to mental health.

It is apparent that Wittenborn's results question the adequacy of certain standard Rorschach scoring categories, particularly those of movement, vista, texture, and the color variables. Since these results, if confirmed, have considerable importance for future Rorschach validation, it seemed desirable to conduct a similarly oriented analysis involving a different psychiatric population. The present study, the first unit of a research program involving factorial analysis of the Rorschach, Wechsler-Bellevue, and the Minnesota Multiphasic Personality Inventory, was designed to do this. In addition, in this study, Wechsler-Bellevue Verbal and Performance IQ scores were included to discover whether there are Rorschach determinants which cluster with a standard measure of intelligence to form a separate factor. Rorschach literature [1, 2, 3] has suggested that an intelligence factor should be most heavily loaded with *W, M, R* and accurate form level.

Procedure

Rorschach protocols and Wechsler-Bellevue scores of 86 psychiatric patients were selected at random from the files of the Clinical Psychology Section of Letterman Army Hospital. These protocols were scored in accordance with procedure advocated by Klopfer [1]. The Rorschach scores as well as the verbal and performance intelligence quotients were intercorrelated, using tetrachoric coefficients.[1] The resulting correlation matrix was factored by the Thurstone centroid method. After extraction of the fifth factor the residuals were small and there was no evidence of common factor variance remaining. The centroid factors were then submitted to a series of orthogonal rotations using the criteria for simple structure. The correlation matrix, the centroid factor loadings, and the rotated factor loadings are shown in Tables 1, 2, and 3. The clarity of these factors might be improved by oblique rotation since loadings would then be higher and prediction from the scores improved.

Analysis[2]

Factor I′ appears related to productivity and pure form, containing high loadings on *R, F,*

---

[1] Tetrachoric r's were used since the short range of many of the Rorschach scoring determinants made the more conventional Pearson r's undesirable.

[2] Because of the magnitude of the standard error of tetrachoric coefficients and the relatively small N in the present sample, it was decided to set a lower limit of approximately .40 for a loading to be considered significant.

Table 1

Intercorrelation among 20 Rorschach Determinants and the Wechsler Bellevue Verbal and Performance IQ's for 85 Psychiatric Patients

Symbol	Variable No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
W	1																					
D	2	19																				
d	3	17	80																			
Dd	4	44	55	41																		
S	5	25	25	01	47																	
M	6	37	75	32	67	52																
FM	7	31	46	23	56	00	50															
m	8	39	31	31	64	19	43	58														
k	9	29	29	35	12	04	12	25	25													
K	10	69	24	21	26	15	33	20	60	25												
FK	11	24	22	27	09	38	29	21	20	09	04											
F	12	32	84	80	60	34	47	16	17	34	–02	14										
Fc	13	28	25	10	24	25	26	22	32	32	10	43	00									
c	14	55	48	37	28	–26	25	26	21	30	24	08	46	33								
C'	15	44	19	31	32	–05	10	16	18	16	11	04	03	13	37							
FC	16	43	40	47	58	21	47	50	55	41	30	09	41	40	35	30						
CF	17	55	21	20	40	02	40	29	10	30	33	19	17	37	56	46	40					
C	18	43	09	00	21	–20	13	–16	–05	26	30	–05	12	01	45	31	09	44				
P	19	10	08	23	14	–06	03	45	24	–27	–25	–05	08	25	–01	24	05	–09	17			
R	20	51	92	69	77	34	65	55	40	65	22	41	90	41	59	28	61	49	06	05		
Vb IQ	21	69	25	18	31	00	30	10	00	32	29	04	15	12	44	12	30	22	22	–10	51	
Pf IQ	22	57	48	36	42	31	30	–08	–13	67	35	07	39	20	24	12	30	25	45	15	56	82

Table 2
Centroid Factors

		I	II	III	IV	V	h^2
1	*W*	76	−44	25	−21	−20	91
2	*D*	77	39	−33	28	07	93
3	*d*	61	26	−20	26	31	65
4	*Dd*	76	28	09	03	−31	76
5	*S*	27	31	−11	−41	−16	38
6	*M*	66	27	09	02	−18	55
7	*FM*	51	41	37	06	−16	60
8	*m*	51	33	51	05	−33	74
9	*k*	51	−25	−22	15	09	40
10	*K*	45	−30	26	06	−41	53
11	*FK*	28	22	08	−32	30	33
12	*F*	66	34	−47	28	19	90
13	*Fc*	44	10	24	−36	29	47
14	*c*	60	−27	13	33	29	64
15	*C′*	39	−18	29	10	23	32
16	*FC*	66	14	15	10	−11	50
17	*CF*	55	−23	37	22	18	57
18	*C*	29	−54	25	15	26	52
19	*P*	14	24	22	−23	17	21
20	*R*	97	26	−31	15	10	1.13*
21	*Vb* IQ	57	−53	−29	−23	−24	79
22	*Pf* IQ	69	−45	−53	−33	00	1.08

* Communalities above 1.00 are probably a result of the small *N* in this sample and the large standard error of tetrachoric *r*.

Table 3
Rotated Factor Matrix

		I′	II′	III′	IV′	V′	h^2
1	*W*	−30	82	21	14	29	90
2	*D*	68	47	26	35	24	92
3	*d*	64	33	−02	00	34	64
4	*Dd*	31	53	58	01	20	76
5	*S*	19	24	29	37	−25	38
6	*M*	30	44	47	06	21	54
7	*FM*	22	15	60	14	37	59
8	*m*	05	18	74	07	39	73
9	*k*	18	54	−16	−15	17	40
10	*K*	−31	51	35	−12	20	53
11	*FK*	21	12	04	51	14	33
12	*F*	82	42	01	−11	15	89
13	*Fc*	07	26	08	56	29	47
14	*c*	60	−25	21	06	42	64
15	*C′*	−05	27	−03	09	50	33
16	*FC*	22	45	39	02	34	52
17	*CF*	−05	36	03	−02	66	57
18	*C*	−29	33	−29	−06	49	51
19	*P*	08	−04	15	39	16	21
20	*R*	75	69	18	01	27	1.14
21	*Vb* IQ	−15	86	−08	−05	−15	77
22	*Pf* IQ	10	96	−28	13	−16	1.06

D, d, and *c*. The highest loadings in Factor III′ involve movement variables in addition to *Dd* and *FC*. Lack of "perceptual control" [5, p. 471] seems to characterize Factor V′. The finding of this factor in the present sample confirms Wittenborn's impression that there are consistent individual differences in the degree to which subjects are spontaneous or uncontrolled in their perceptual approach.

Few significant loadings are found on Factor IV′, which forces a conservative interpretation. It should be noted, however, that the loadings in the shading variables, *FK* and *Fc*, are consistent with Rorschach scoring procedures, i.e., scoring shading separately from other variables. Factor II′ is of interest in that it supports the clinical impression that the Rorschach can be used as a predictor of intelligence, a finding which has already received support in previous research [2, 3]. It contains variables which have been utilized by Rorschach interpreters to evaluate intelligence, namely, *W*, *M*, *F*, *FC*, and*R*. The appearance of diffuse shading (*K*, *k*) in this factor is perhaps not surprising, since it undoubtedly requires higher than average intelligence to verbalize such percepts.

Comparison

The five factors described above provide additional support for Wittenborn's four hypotheses. Factors I′, III′, and V′ are similar in composition to three of the four factors obtained by that author in his psychiatric sample. Except for *W* and *K*, the fifth factor in this analysis is similar to his "perceptual control" factor. In the present investigation variables *W* and *K* appear on Factor II′, "the intellectual factor." Factor III′ as in his study involves the clustering of *M*, *FM*, and *FC*. However the appearance of *m* in this factor provides some evidence of a separate "movement" dimension. Factor I′ in this sample is similar to the "productivity" dimension found in his psychiatric sample. The presence of *c* in this factor is not clearly anticipated in the previous studies. However, the factorial composition of *c* here obtained is similar to that found in his second study. The factorial pattern of *FC* in the present analysis is also much more similar to that of *M* than to *CF* and *C*. The appearance of *FC* in conjunction with the movement scores further confirms Wittenborn's concept that *FC* is different from the

CF and *C* determinants and is closely allied with *m, M,* and *FM* in its behavioral implications.

The shading variables *FK* and *Fc* form a separate factor. Wittenborn did not obtain such a dimension among his scores, and concluded that shading should not be scored and interpreted independently of other variables such as color. Factor IV′, then, provides some support for the common practice of scoring shading separately from other variables, although it does not demonstrate that fine distinctions drawn between the various qualities of shading are warranted. Wittenborn's analysis also generated a factor composed primarily of *S* and *O,* which he suggested was related to level of mental health. Since *O* was not scored in this set of protocols, it is not possible to confirm these findings.

Summary

The present analysis was undertaken for two reasons: (*a*) to replicate an earlier study by Wittenborn on a different population, and (*b*) to discover whether there is an intellectual factor among the several scores in a standard Rorschach scoring system.

Twenty Rorschach determinants and two Wechsler-Bellevue intelligence scores from the protocols of psychiatric patients of Letterman Army Hospital were intercorrelated and the resulting matrix analyzed by the Thurstone centroid method. In the analysis five factors are necessary to account for the common factor variance as compared to four in Wittenborn's two previous analyses. The five factors appear to represent the following dimensions: (*a*) "productivity," (*b*) "movement," (*c*) "lack of perceptual control," (*d*) "shading," and (*e*) "intelligence."

Wittenborn's original four hypotheses were confirmed, as was his hypothesis that *FC* was more closely allied to movement than to color determinants. In addition, the appearance of an "intelligence" factor supports the belief that certain Rorschach determinants covary with intelligence.

Received March 25, 1953.
Early Publication.

References

1. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N.Y.: World Book Co., 1942.
2. Tucker, J. Rorschach human and other movement responses in relation to intelligence. *J. consult. Psychol.,* 1950, **14**, 283–286.
3. Wishner, J. Rorschach intellectual factors in neurotics. *Amer. J. Orthopsychiat.,* 1948, **18**, 265–279.
4. Wittenborn, J. R. A factor analysis of Rorschach scoring categories. *J. consult. Psychol.,* 1950, **14**, 261–267.
5. Wittenborn, J. R. Level of mental health as a factor in the implications of Rorschach scores. *J. consult. Psychol.,* 1950, **14**, 469–472.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

The Effect of Test Anxiety on Memory Span in a Group Test Situation

Donald Kaye, Paul Kirschner, and George Mandler
Yale University

The effect of anxiety on test performance has been investigated in several studies in the literature. Generally, this effect has been demonstrated under the condition of the experimental introduction of stress. Williams [6] found the stress effect in the performance of the Digit-Symbol substitution test and Moldawsky and Moldawsky [2] showed a decrement on the Digit Span subtest of the Wechsler-Bellevue scale [5].

In several previous studies [1, 3, 4] an attempt has been made to show the effect of test anxiety level on the individual performance of typical intelligence tests. Briefly, a questionnaire designed to measure specific test anxiety was administered to a large group of subjects. Two ends — Low Anxiety and High Anxiety subjects — of the resulting distribution were then tested. Differential results for the two groups were obtained on the Kohs Block Design test [1], the Digit-Symbol test [4], and the Mathematical Aptitude test [3].

In the study presented here, the effect of test anxiety on memory span in a group test situation was investigated. Moldawsky and Moldawsky [2] have established the effect of situational stress on the Digit Span test in an individual test situation. This study supplements their procedure in examining the diagnostic value of a memory span test in differentiating two groups of *S*s whose anxiety level has been independently measured.

Procedure

Thirty-six students in introductory psychology courses, who had previously taken the anxiety questionnaire, described in detail elsewhere [3], served as *S*s. Seventeen *S*s were taken from the lower 13% of the distribution of anxiety scores and nineteen *S*s represented the upper 13% of the distribution. None of the *S*s knew of any connection between the questionnaire and the testing session.

The 36 *S*s were tested, for purposes of convenience, in three groups of 8, 14, and 14 respectively. At the beginning of a group session the *S*s were told that the tests they were about to take measured the memory component of the IQ and that their scores would be available afterwards. These instructions were given in line with previous findings [4] that only under these, presumably anxiety-evoking, conditions did a differential effect between the groups appear.

The memory span test lasted twenty minutes and consisted of 28 items: 10 items were number series varying in length from six to ten; 10 items consisted of mixed letters and numbers with the same variation in length and 8 items contained lists of short words (e.g., number, nail, white, foot, light, soot) varying from six to nine words per list. The *S*s were required to write down as much of an item as they could recall immediately after it was read to them. The correlation between the memory span test and an alternate form was .82 (Pearson r) for the 36 *S*s.

Results

The tests were scored for the number of correct reproductions. The mean score of the Low Anxiety group was 142.4, for the High Anxiety group 127.9; the t value of 2.37 is significant at the .03 level of confidence.

This result bears out two main points:

1. A memory span test significantly differentiated two groups with differential levels of test anxiety as measured by an anxiety questionnaire. The Low Anxiety group performed significantly better than the High Anxiety group.
2. The differential effect of anxiety levels

operates in group situations as well as individual testing situations when ego-involving instructions are given.

Received October 28, 1952.

References

1. Mandler, G., & Sarason, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.
2. Moldawsky, S., & Moldawsky, P. C. Digit span as an anxiety indicator. *J. consult. Psychol.*, 1952, **16**, 115–118.
3. Sarason, S. B., & Mandler, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 810–817.
4. Sarason, S. B., Mandler, G., & Craighill, P. C. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 561–565.
5. Wechsler, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.
6. Williams, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1947, **11**, 21–29.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Neurocirculatory Reactions and a Group Stress Situation

George B. Strother and Donald M. Cook[1]
University of Minnesota, Duluth Branch

Abel [1] found that differences in the recall of incompleted and completed tasks by adolescents were related to differences in the Schneider index of autonomic efficiency. Her findings showed that subjects with good autonomic function as indicated by Schneider indices of 11 or higher recalled more incompleted tasks and fewer completed tasks, whereas subjects with low indices (0 - 6) recalled more completed tasks. Stagner [4] interpreted these findings as indicating that the low-index group responds to interruption as a failure situation leading to repressive forgetting, whereas high-index subjects respond to interrupted tasks as a challenge and tend to recall more of them than of the completed tasks.

The purpose of this study was to perform a parallel experiment using an obviously frustrating test situation instead of the Zeigarnik experiment.

The frustrating test situation selected was McKinney's "reporting test" [2] described below. Since it was demonstrated by McKinney that this situation produced marked individual differences in response to the experimental stress, the problem was to determine whether these differences were related to the subjects' Schneider index.

Procedure

The subjects. The subjects were 46 college students enrolled in an introductory course in psychology.

The "index of autonomic efficiency." The Schneider index [3] is designed to determine the efficiency of autonomic regulation of mild cardiovascular stress by assigning point values to pulse rate and blood pressure readings and changes, when reclining and standing, and at intervals following exercise.

The reporting test. The test is divided into three parts: in Part 1 subjects work for 8 minutes without interruption reporting the numbers of each of three different kinds of objects contained in a series of circles. Then for 16 minutes they are compelled to perform the same task against impossible time limits. They are frequently interrupted and told to mark "failure" (F) on that portion of their work. Finally, they work again for 8 minutes without interruption. The scores employed in this study were those recommended by McKinney as being most significant—the differences between the percentage correct in Parts I and III, respectively.

Results

Table 1 has been arranged to parallel, as nearly as possible, the summary of Abel's data as presented in Stagner [4, p. 327]. It can readily be seen by comparing the two tables that the results are in striking agreement. The majority of high-Schneider subjects in Abel's

Table 1
Autonomic Function and Response to a Frustrating Situation

Schneider index	Reporting-test reaction, percentage of cases showing: Marked increase in errors (-38 to -9)	Moderate increase in errors (-8 to -1)	No increase or decrease in errors (0 to 10)	N
High (11-17)	32	41	27	22
Medium (7-10)	41	41	18	17
Low (0-6)	72	28	0	7

---

[1] The authors are indebted to Dr. Fred McKinney of the University of Missouri for the use of the "reporting test" materials.

study remember more interrupted tasks; in the present study more high-Schneider subjects respond to the reporting-test situation by showing no loss or actual gain from Condition I to Condition III. Conversely, in Abel's study the majority of low-Schneider subjects remember more completed tasks; in the present study, the majority of low-Schneider subjects show substantial loss under Condition III.

Table 1 was planned to permit direct comparison with Abel's results. As one consequence, the number of cases in the low-Schneider group is small. The present group of subjects, possibly because of age or some other factor, tended to have a higher average than Abel's adolescents. In order to test whether the results so presented did represent a genuine relationship, a product-moment *r* between reporting-test scores and Schneider indices was computed. Table 2 gives the means, standard deviations, and the intercorrelation of the reporting test and the Schneider index. The coefficient of .30 is significant at the .05 level.

Table 2

Relationship between Schneider Index and the Reporting Test

Variable	Mean	*SD*	*r*	*p*
Reporting test	−8.75	10.09	.30	.05
Schneider index	10.46	3.38		

Discussion and Summary

Numerous physiological factors have been suggested as having some relationship to various personality variables. The autonomic nervous system has been the subject of a fair share of such speculation and investigation. It has seemed plausible to assume that such characteristics as emotionality, nervousness, excitability, or lability might be related to individual differences in the excitability of the autonomic nervous system. Much of the theoretical and experimental work dealing with the concept of frustration tolerance seems to assume the existence of a generalized tendency for individuals to react characteristically to a variety of threatening or frustrating situations. Such a generalized tendency might well be at least partially the result of individual differences in autonomic excitability or efficiency. Abel's work seems to be in accord with this hypothesis if we assume that interrupted tasks appear threatening or frustrating to some subjects.

In order to confirm Abel's finding, using a task which is specifically designed to be disturbing to the subjects, these experimenters used McKinney's reporting test on a group of college students whose Schneider indices they had previously determined. The results accord closely with Abel's and indicate that there is an apparently significant relationship between an individual's autonomic reactions and his response to experimental frustration. The Schneider index measures the readiness with which the individual makes a neurocirculatory adaptation to moderate physiological stress; the reporting test measures the individual's ability to adapt behaviorally to psychological stress. There seems to be a measurable degree of relationship between the two.

Received November 25, 1952.

References

1. Abel, Theodora M. Neurocirculatory reactions and the recall of unfinished and completed tasks. *J. Psychol.*, 1938, **6**, 377–383.
2. McKinney, F., Strother, G. B., Hines, Ruth R., & Allee, Ruth A. Experimental frustration in a group test situation. *J. abnorm. soc. Psychol.*, 1951, **46**, 316–323.
3. Schneider, E. C. *The physiology of muscular activity.* Philadelphia: Saunders, 1933.
4. Stagner, R. *Psychology of personality.* (2nd Ed.) New York: McGraw-Hill, 1948.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Responses of Women to the Thematic Apperception Test

Leonard D. Eron
Yale University

For meaningful interpretations of TAT productions reference should be made to normative data obtained in a standard fashion from well-defined segments of the population. A study by Ritter and Eron [9] has demonstrated that fantasy is important for personality evaluation primarily when it is deviant. This is not to say that fantasy content which is common to a majority of individuals in a given age and cultural group is not important for the understanding of the personality of these individuals. However, in order to diagnose an individual as primarily maladjusted on the basis of projective test responses, account must be taken chiefly of the ways in which responses differ from those of others in the same age and cultural group. It is important that the experimenter not be misled in his interpretations by responses which seem unusual to the "naked eye" but which on more careful investigation are found to be common among subjects of similar age and status.

The results reported here are based on an extension to two groups of female subjects of a methodology for analysis of TAT stories which was developed with some success on male subjects [3, 4]. It is a procedure for noting and summarizing all the important data in a TAT protocol so that it can easily be compared to normative material and significant deviations from expected responses can then be indicated. This can be done both for responses to individual cards by use of what Rosenzweig has termed "apperceptive norms" [10] and for consistencies throughout all 20 cards in a given protocol by using "thematic" norms. Both types of norms are here presented.

Method

Subjects. The subjects for this investigation included two groups of female subjects: Group C, 40 female college students enrolled in a course in introductory psychology at the University of Wisconsin;[1] Group P, 20 women in the third to fifth month of pregnancy and attending the Obstetrical Clinic of the New Haven Hospital.[2] The age range of Group C was 18 to 31 with a median of 23 years. The age range of Group P was 19 to 35 with a median of 24. All the women in group P had IQ scores of at least 100 on the Wechsler-Bellevue Scale; 6 had a college education, 6 high school, and 8 grammar school only; none had ever seen a psychiatrist except in the current investigation, and none was judged by the psychiatrist to have severely debilitating emotional problems beyond what would be expected in a randomly selected normal group. They represented a cultural cross section of the New Haven population. None of the subjects in Group C had ever been or were at that time under psychiatric treatment. Since they were college students, it is assumed their intelligence was at least normal. Culturally they probably represent a more highly selected group than the subjects in Group P, although they did come from all sections of the country.

Procedure. The Thematic Apperception Test was individually administered to all subjects in a standard manner described elsewhere [3]. All 20 cards recommended for adult females by the authors of the test [8] were included.

Analysis of data. Each story was analyzed for emotional tone, outcome, themes, identification of characters and objects, perceptual distortions and unusual formal characteristics.

---

[1] These subjects were included in an investigation by Dr. Dorothy Terry [11], to whom thanks are expressed for use of these protocols.

[2] These subjects were being studied in connection with research on natural childbirth [6] in which the author participated and which was under the direction of Dr. Lawrence Freedman.

Only one variable was rated or scored at a time and all the stories for a given picture were rated together.

The scales for emotional tone and outcome are the same as those used in previous studies where they were described, their construction was discussed, and their reliability was established [3, 5]. A further check was made on reliability of the scales in this study since different individuals made the ratings.[3] Two-hundred stories, selected randomly from protocols of all the subjects, were rated by two psychologists for emotional tone and outcome. The Pearson r between raters for emotional tone was .86 and for outcome was .75. These coefficients are comparable to those established in the original studies and it would seem thus that this is the order of reliability of these scales, regardless of the individuals making the ratings. It was necessary to add an additional category, x, to the rating scale for outcome to cover those stories in which no outcome was given, or which were mere descriptions with no action taking place. In the original study there was no provision for such a category since the subjects were always asked to supply an outcome for each story whenever it was omitted.

An attempt was made to use directly the thematic check list devised for the original study with males [2]. It soon became apparent that the identical list could not be used for both sexes. In the first place, eight of the cards recommended for females are different from those used with male subjects, and new themes had to be introduced; also some of the old themes were not applicable. More important than this, though, was the fact that the original list had been drawn up from the point of view of a male hero since it was assumed that the male subjects would identify with the male characters in the story most like himself in age, etc., as suggested by the authors of the test [8]. Logically, this list could not then be used with female subjects if it is assumed that in general they would in the same way identify with female characters. However, for many of the cards it seemed that the central characters in the stories told by a majority of the female subjects were males. This was especially true for cards 4 and 13 MF, both of which contain a male and a female figure. To check on this, two female experimenters went through all the responses to these cards and made judgments as to which character in each story seemed to be the central one from whose point of view the story took shape. Percentage agreement as to choice of central character was 94 and each examiner selected the female character as central in 26% and 28% of the stories for card 4, and in 4% and 6% of the stories for card 13MF. This result is similar to that obtained in an unpublished study by Cohen [1] in which 50 women subjects were asked who the central character was in their stories. For card 4 he found that 80% of these female subjects chose the male, 20% the female, and for card 13MF, 86% chose the male and 14% the female. Thus it seemed that with female subjects a check list had to be used which was not drawn up from the point of view of a hero or heroine, at least insofar as the heterosexual situation was concerned. Consequently the old check list was revised and Table 1 sets forth this new check list for thematic analysis.

Table 1

Check List of Themes for Analysis of Women's TAT Stories

I. Disequilibrium (tension)

A. *Interpersonal*

1. Parent
   - *a.* *Pressure*—parent or parent figures are prohibitive, compelling, censuring, punishing, disapproving, interfering, checking up, disagreeing with, quarreling with, restraining or unduly influencing child.
   - *aa.* Child is exerting pressure on parents.
   - *b.* *Succorance*—child seeks or receives aid, help, advice, consolation from parent.
   - *c.* *Nurturance*—child bestows or offers advice, aid, consolation to the parent.
   - *d.* *Aggression from*—physical harm inflicted or intended upon child by parent.
   - *e.* *Aggression to*—physical harm inflicted or intended upon parent by child.
   - *f.* *Departure*—child is taking leave of parental home, runs away from home.
   - *g.* *Concern*—parent is worried over physical or mental well-being of child.

---

[3] Mrs. Ruth Straatsma rated all the stories for emotional tone, and Mrs. Margaret Ingram rated those stories used in the reliability estimate. Mrs. Sophie Dutton rated all the stories for outcome, and reliability was checked with ratings made by the author. The author analyzed all stories for themes, unusual formal characteristics, and perceptual distortions.

gg. *Concern*—child is concerned over parent.
h. *Incest*—actual or contemplated.
i. *Death or illness* of parent.
j. *Death or illness* of child.
k. *Disappointment to*—parent is disappointed in child's behavior or accomplishments. Parent rejects child.
l. *Disappointment in*—child is disappointed by parent, ashamed of parents, derogates parent, rejects parent.
m. *Filial obligation*—child feels it is his duty to remain with, comply with, or support parents.
n. Confession—child tells parent of some misdemeanor or crime he has committed.
nn. *Confession*—parent confesses to child.
o. *Bad news*—child brings sad tidings to parent (e.g., death of other parent, brother, etc.).
oo. *Bad news*—parent brings sad tidings to child.
p. *Marriage*—child tells parent of past or impending marriage, parent objects to marriage, interferes with marriage.
q. *Collusion*—parent and child are planning or executing together some antisocial act.
r. *Parental conflict*—child is concerned over marital problems of parents. Parents disagree over child care, etc.
s. *Lost*—child has been abandoned by parents or is lost.
v. Facts of life explained to child by parent.
w. Parent tells child a sibling is coming.
x. Parent is jealous of child.

2. Partner (husband, sweetheart, opposite sex peer)
a. *Pressure*—heroine and partner are quarreling, can't agree.
b. *Nurturance*—male character bestows or offers to partner aid, consolation, advice, etc., or these are being sought by female.
c. *Succorance*—male character seeks or receives aid, comfort, consolation, assistance from partner, advice from partner.
d. *Aggression from female character*—physical harm inflicted upon or intended for partner by female.
e. *Aggression from male character*—physical harm inflicted upon or intended for partner by male.
f. *Departure from*—male character leaves partner either temporarily or permanently.
ff. *Departure from*—female character leaves partner either temporarily or permanently.
g. *Concern*—male character is worried over physical or mental well-being of partner.
gg. *Concern*—female character is worried over physical or mental well-being of partner.
h. *Illicit sex*—extra- or premarital heterosexual relationship, nonincestuous, includes "petting."
i. *Illicit sex with violence*—rape.
j. *Death or illness* of female partner.
jj. *Death or illness* of male partner.
k. *Disappointment*—male character is disappointed in female's behavior, or accomplishments, rejects female, derogates female.
l. *Disappointment*—female character is disappointed in male's behavior, or accomplishments, derogates male, rejects male.
m. *Jealousy*—female character is jealous of partner's attention to others or other's attentions to partner.
n. *Competition*—male character is jealous of partner's attention to others or other's attention to partner.
o. *Unfaithful*—male character discovers partner has been having extramarital relations or finds her raped.
oo. *Unfaithful*—female character discovers partner has been unfaithful.
p. *Decision*—male character must choose between marriage and not or between two partners.
pp. *Decision*—female character must choose between marriage and not or between two partners.
q. *Pursuit*—male character is wooing or trying to get partner to submit, wants contact with other sex.
r. *Seduction*—female character is actively pursuing or talking male into relationship, wants contact with other sex.
s. *Unrequited*—male character's love is unreturned, impotence. Waits for partner.
ss. *Unrequited*—female character's love is unreturned, is "done in" by partner. Waits for partner.
t. *Restraint*—central character and partner can't be married or have intercourse (because of economics, deformities, menstrual periods, etc.).
u. *Childbirth*—pregnancy.
y. *Surprise*—female character surprised by partner.

3. Peer
a. *Pressure*—like-sexed peers are prohibitive, compelling, censuring, punishing, disapproving, interfering, restraining, etc.
b. *Succorance*—central character seeks or receives aid, advice, consolation from peer.
c. *Nurturance*—central character bestows or offers aid, advice, consolation upon peer.
d. *Aggression from*—physical harm inflicted upon or intended for central character by peer.
e. *Aggression to*—physical harm inflicted upon or intended for peer by central character.
f. *Desertion*—central character has been deserted by his companions or is lost from them.
g. *Concern*—central character is worried about physical or mental well-being of friends.
h. *Homosexuality*—actual or contemplated.

   i. *Death or illness* of friend.
   j. *Belongingness*—desire expressed to be with or accepted by peers.
   k. *Unappreciated by peers.* Rejected by peers, peers not interested in welfare.
   l. *Competition*—central character is competing with peer in game or contest, or professionally.
   m. *Hypnotism.*
   n. *Envy*—central character is envious of peer's accomplishments, capacities, belongings, etc.
   nn. Others envy central character.
   o. *Revenge*—central character is anxious to extract revenge from a peer for some past deed or is taking the revenge.
   p. *Disregard for*—central character is unconcerned over welfare of peers, deserts peers.
   q. *Devaluation of peer*—derogation, rejection of peer by central character.

4. Sibling
   a. *Pressure*—siblings are prohibitive, compelling, censuring, punishing, disapproving, interfering, restraining, etc.
   b. *Succorance*—central character seeks or receives aid, advice, consolation from sibling.
   c. *Nurturance*—central character bestows or offers aid, advice, consolation upon sibling.
   d. *Aggression from*—physical harm inflicted upon or intended for central character by sibling.
   e. *Aggression to*—physical harm inflicted upon or intended for sibling by central character.
   f. *Rivalry*—siblings are striving for same end, are jealous of each other, etc.
   g. *Concern*—central character is worried about physical or mental well-being of sibling.
   h. *Incest*—actual or contemplated.
   i. *Death or illness* of sibling.

B. *Intrapersonal*

1. *Aspiration*—dreaming of future, hoping for future, determination.
2. *Inadequacy*—realization, whether justified or not, of lack of success; individual is at a loss to cope with situation.
   a. *Unrealized.*
3. *Curiosity*—wondering about construction of object, contents of room, etc., desire to observe, inquire, explore, investigate; to acquire facts.
4. *Behavior disorder*—personal maladjustment of all sorts, neurosis, psychosis, sleeplessness, hallucination, withdrawal, eccentricity, morbid preoccupation, nightmare, excessive emotionality.
5. *Suicide*—attempted or complete, preoccupation with.
6. *Moral struggle*—concern over what is right and wrong, hesitancy in indulging in some act because of ethical proscriptions.
7. *Guilt*—remorse, shame.
8. *Drunkenness.*
9. *Fear-worry*—central character is apprehensive, alarmed, terrified, of some person, thing, force.
10. *Rumination*—inability to understand world and its problems, adolescent reverie, man's insignificance, "what am I?", "what's it all about?".
11. *Occupational concern*—deciding between jobs, considering vocations, dissatisfied with present employment or present pastime; worried about job. Emphasis on decision which must be made in choosing occupation.
12. *Physical illness or death of central character*—(other than when in a peer, partner, sibling, parent relationship, which are included under those specific categories).
13. *Retribution*—forced to atone or be punished for some antisocial act.
14. *Reminiscence, sad*—individual is unhappy in his memories of the past or contemplation of the future, hopelessness.
15. *Intra-aggression*—does physical harm to self short of suicide, (which is carried under that heading), either through accident or deliberate, looks forward to own death.
16. *Religion*—prayer, seeking consolation from God, religious conflict, religious awakening.
17. *Loneliness*—central character misses someone, is an outcast, friendless, homeless.
18. *Compensation*—when individual has one characteristic or stroke of fortune to make up for another bad characteristic or misfortune.
19. *Vacillation*—wasting time, putting off a distasteful task, procrastination, loitering, waiting for someone unspecified.
20. *Acquisition*—desire expressed by central character to acquire material things, or is working greedily for possession of goods; miserliness.
21. *Exhaustion*—characters completely "pooped" from overexertion.
22. *Revenge*—central character is preoccupied with wish to retaliate for some past wrong from unspecified individual (when peer, carried under that heading).
23. *Inconsequential sadness*—over death of dog, broken toy, loss of object, etc.
24. *"Hurt Feelings"*—central character is very sensitive over some happening, slight, insult, etc.
25. *Jealousy*—of unspecified individual (if specific person named, carried under appropriate heading).
26. *Self-pity*—central character feels sorry for self.
27. *Supernatural*—return of the dead, ghosts.
28. *Homesickness*—expressed by central character.
29. *Physical illness or death* of unspecified character (not central character, partner, peer, sibling).
30. *Bored*—central character is bored, has nothing to do.
31. *Regression*—central character is unwilling to grow up, assume adult responsibilities, grow old.
32. *Posing*—central character posing for portrait.
33. *Accusation*—central character accused, usually falsely, of some misdemeanor, crime, infidelity.

C. *Impersonal*

1. *Economic pressure*—individual is compelled to, prohibited from, or limited in doing something because of the lack of money.
2. *Legal restriction*—individual is incarcerated, arrested, or detained against his will.
3. *Generalized restriction*—environment is generally frustrating because of backwardness, danger, lack of opportunity, etc.; hopelessness of life.
4. *Aggression towards environment*—robbery, accident, murder of unspecified individual (if individual named, carried under the proper heading).
5. *Aggression from*—impersonal source, accident, animal, nature, disease, etc.
6. *War*
7. *Escape from perilous (restricting, shameful) environment*—individual is in the act of getting out (doesn't include rescue, which comes under II).
8. *Obligation to society*
9. *Crime without punishment*

II. Equilibrium

A. *Interpersonal*

1. Parent
    - *a. Cooperation*—parent is working with child towards his own goal.
    - *b. Resignation*—parent is resigned to child's activity.
    - *c. Idealization*—child idealizes parent, wants to be like parent.
    - *d. Reunion*—of child with parents.
    - *e. Fulfillment*—child lives up to expectations of parents.
    - *f. Contentment*—blissful home situation, stress security of home.
    - *g. Ordinary familial activity*—just talking, spending an evening, working, etc.
2. Partner
    - *a. Admiration*
    - *b. Cooperation*—between mates, so that both are working for the same end and are happy.
    - *c. Contentment*—mutual love and affection.
    - *d. Reunion*
    - *e. Ordinary activity*
    - *f. Proposal*
    - *g. Anniversary*
    - *h. Dancing*
3. Peer
    - *a. Cooperation*—working together for same goal.
    - *b. Congeniality*—sociability, friendliness, satisfactory affiliation.
    - *c. Reunion*—with friends.
    - *d. Approbation*—central character is lauded, appreciated by friends, and peers.
    - *e. Exhibition*—central character is amusing, entertaining, attracting others.
    - *f. Ordinary activity*—everyday association with nothing in particular happening.
4. Sibling (no II themes found in original sample of 200 cases).

B. *Intrapersonal*

1. *Self-esteem*—confidence, belief in own superiority, conceit.
2. *Tranquility*—peace of mind, content with environment and own accomplishments, enjoyment, aesthetic appreciation.
3. *Reminiscence, happy*—individual is content with her memories, or contemplation of future.
4. Retirement—central character is asleep, resting, no physical exhaustion, relaxation.
5. *Occupational satisfaction.*
6. *Resignation to lot*—individual accepts situation without further struggle.
7. *Ordinary activity*—individual is going about her everyday activities.

C. *Impersonal*

1. *Favorable environment*—individual is being helped by favorable circumstances, is enjoying his surroundings.
2. *Rescue*—successful escape from perilous environment or cessation of noxious stimuli (already accomplished).

The list of unusual formal characteristics is the same as that used in the former studies [3, 9] under the heading of levels of interpretation and includes such things as autobiographical stories, symbolic or abstract stories, descriptions rather than narratives, denial of popular themes, rejections and alternate plots. Perceptual distortions pertain to any marked deviations from the descriptions of objects and characters furnished by the authors in the manual of instruction [8].

Results

Analysis in these terms (emotional tone, outcome, themes, formal characteristics, and perceptual distortions) are presented for stories in response to each card and also for all 20 stories in a protocol, so that there is information concerning what are the common apperceptive responses for each card and also an indication of how many individuals have a given number or more of each kind of response in a total of 20 stories. In the apperceptive norms, the data are presented in terms of percentage of subjects having a specific response for any one card, and in the thematic norms, in terms of the percentage of subjects having a given number or more of these responses in a total of 20 stories. Thus, by referring to Table 2, for example, it can be seen that 52% of the subjects gave stories with emotional tone rating of —1 to card 1; and by referring to Table 7, it becomes apparent that 75% of the subjects contributed seven or more stories with emotional tone rating of —1 in a total series of 20 stories. Significant differences beyond the .05 level of confidence between the two samples of subjects were noted. In calculating these differences the actual frequencies rather than percentage frequencies were used in the standard formula for calculating the chi square of homogeneity, with Yates's correction where appropriate [7]. Fourfold tables, denoting presence or absence for each group, were utilized in calculating differences in themes, unusual formal characteristics and perceptual distortions. With emotional tone and outcome, where 5 × 2 tables were possible because of rating-scale intervals, appropriate cut-off points were used so that there would be a theoretical frequency of at least five in each cell. Consequently there were a number of 3×2 as well as 2 × 2 tables. In the thematic norms, which indicate the percentage of individuals having a given number or more of any response in a total of 20 stories, the chi squares were calculated from the actual frequency of individuals having each number of instances in question. Cut-off points here were also selected so that the theoretical frequency of any cell would be at least five and there were thus a number of 3 × 2 as well as 2 × 2 tables.

Table 2

Percentage Frequency of Subjects Contributing Stories of Various Emotional Tone Rating for Each of 20 Cards

Card	*Emotional Tone Ratings*					
	—2	—1	0	+1	+2	Reject
1	0	52	22	22	5	0
2	0	48	37	15	0	0
3GF	38	62	0	0	0	0
4	12	87	0	0	0	0
5	3	45	48	3	0	0
6GF	47	28	25	0	0	0
7GF	5	23	50	22	0	0
8GF	22	62	17	0	0	0
9GF	13	72	12	2	0	2
10	13	27	3	43	13	0
11	13	57	30	0	0	0
12F	27	23	37	13	0	0
13MF	77	22	2	0	0	0
14	3	13	40	42	2	0
15	47	42	12	0	0	0
16	2	13	35	37	10	3
17GF	12	27	43	17	0	2
18GF	43	52	3	2	0	0
19	5	27	27	40	0	2
20	7	40	52	0	0	2

Table 3

Percentage Frequency of Subjects Contributing Stories of Various Outcome Ratings for Each of 20 Cards

Card	Outcome Ratings						
	—2	—1	0	+1	+2	?	*x*
1	3	25	7	28	18	7	12
2	5	17	17	42	12	5	2
3GF	22	27	5	30	5	5	7
4	3	27	2	35	10	10	13
5	7	15	18	25	7	2	25
6GF	5	15	13	33	5	3	25
7GF	3	13	13	37	8	0	25
8GF	8	2	25	25	8	2	30
9GF	10	22	12	27	3	0	27
10	3	17	8	13	37	2	20
11	18	5	12	5	30	2	28
12F	7	23	10	23	10	3	23
13MF	47	8	3	13	0	8	20
14	2	13	13	40	17	2	13
15	23	3	15	27	3	5	23
16	2	8	15	37	17	0	22
17GF	12	15	18	22	7	7	20
18GF	30	15	0	32	3	5	15
19	3	7	8	27	8	0	47
20	10	10	18	23	12	3	23

Table 4

All Themes Contributed by Two or More Subjects for Each of 20 Cards

Code	Theme	Percentage frequency
Card 1		
IA1*a*	parental pressure	60
IB11	occupational concern	37
IB1	aspiration	35
IA3*j*	belongingness	27
IB2	inadequacy	10
IB7	guilt-remorse	5
IC1	economic pressure	5
IB3	curiosity	3
IB19	vacillation	3
Card 2		
IB1	aspiration	52
IB11	occupational concern	27
IA1*a*	parental pressure	23
IA1*m*	filial obligation	15
IIA1*a*	cooperation with parent	10
IC3	generalized restrictions	8
IA4*f*	sibling rivalry	8
IA1*c*	nurturance to parent	8

IA2*r*	seduction	7
IIA1*g*	ordinary familial activity	7
IC1	economic pressure	7
IB7	guilt	5
IA1*f*	departure from parent	3
IA1*i*	death or illness of parent	3
IIA1*e*	fulfillment, parent	3

Card 3GF

IA1*j*	death or illness of child	15
IA2*a*	pressure from partner	15
IA2*ss*	unrequited female partner	13
IA1*a*	parental pressure	12
IB7	guilt	12
IB29	illness or death of unspecified character	12
IA2*f*	departure from partner	10
IA2*jj*	death or illness of male partner	10
IB17	loneliness	8
IA1*i*	death or illness of parent	8
IB12	physical illness or death of central character	7
IB24	hurt feelings	5
IA1*r*	parental conflict	3
IA2*oo*	unfaithful male partner	3
IB2	inadequacy	3
IB4	behavior disorder	3
IC1	economic pressure	3

Card 4

IA2*c*	succorance from partner	38
IA2*f*	departure from partner	25
IA2*b*	nurturance to partner	18
IA2*ss*	unrequited female partner	13
IA2*p*	decision for male partner	12
IA2*r*	seduction	10
IA2*oo*	unfaithful male partner	8
IB2	inadequacy	7
IB11	occupational concern	7
IC5	aggression from environment	7
IA2*gg*	concern of female partner	5
IA2*h*	illicit sex	5
IB1	aspiration	5
IA2*m*	jealousy	3
IA2*n*	competition	3
IA2*q*	pursuit of partner	3
IA2*s*	unrequited male partner	3
IA3*e*	aggression to peer	3
IB6	moral struggle	3
IB22	revenge	3
IC6	war	3

Card 5

IB3	curiosity	48
IA1*a*	parental pressure	35
IA1*g*	concern of parent	17
IB9	fear, worry	12
IIB7	ordinary activity	8
IA1*k*	disappointment to parent	7
IA2*gg*	concern of female partner	5
IC4	aggression to environment	5
IB4	behavior disorder	5

Card 6GF

IA2*a*	pressure from partner	28
IB9	fear, worry	17
IIA2*e*	ordinary activity	15
IA2*q*	pursuit of partner	10
IA2*r*	seduction	10
IA1*a*	parental pressure	7
IB33	accusation	7
IB7	guilt	5
IC4	aggression to environment	5
IA1*p*	marriage	3
IA2*h*	illicit sex	3
IA2*l*	disappointment in male partner	3
IA2*oo*	unfaithful male partner	3
IA2*y*	surprise by partner	3
IB24	hurt feelings	3
IIA2*d*	reunion with partner	3
IA1*g*	concern of parent	3

Card 7GF

IA1*a*	parental pressure	32
IA1*v*	facts of life	18
IA1*w*	a sibling is coming	12
IA3*j*	belongingness	10
IA4*f*	sibling rivalry	10
IB1	aspiration	8
IB7	guilt	8
IB30	bored	8
IB23	inconsequential sadness	7
IIA1*g*	ordinary familial activity	7
IA1*b*	succorance from parent	5
IA*ll*	disappointment in parent	5
IIA1*f*	contentment with parent	5
IB3	curiosity	3
IB4	behavior disorder	3
IB11	occupational concern	3
IB16	religion	3

Card 8GF

IIB3	reminiscence, happy	33
IB1	aspiration	20
IB11	occupational concern	15
IC1	economic pressure	15
IIB7	ordinary activity	15
IB32	posing	13
IB21	exhaustion	10
IB14	reminiscence	8
IB10	rumination	7
IIB4	retirement	5
IA1*b*	succorance from parent	3
IA1*g*	concern for child	3
IB13	retribution	3
IB17	loneliness	3
IC3	generalized restriction	3
IIB2	tranquility	3

Card 9GF

IC7	escape from perilous environment	33
IB3	curiosity	28
IA2*m*	jealousy	22
IB9	fear, worry	18

IA3*a*	pressure from peer	13
IA2*oo*	unfaithful male partner	13
IA4*f*	sibling rivalry	10
IA2*a*	pressure from partner	7
IA2*e*	aggression from male partner	7
IA2*ss*	unrequited female partner	7
IA3*b*	succorance from peer	7
IC5	aggression from environment	7
IIA3*b*	congeniality with peer	7
IA3*n*	envy of peer	5
IA3*q*	devaluation of peer	5
IA4*a*	pressure from sibling	5
IA2*jj*	death or illness of male partner	3
IA3*c*	nurturance to peer	3
IA3*k*	unappreciated by peer	3
IA3*p*	disregard for peer	3
IB2	inadequacy	3
IB4	behavior disorder	3
IB5	suicide	3

Card 10

IIA2*c*	contentment	40
IA1*j*	death, illness of child	20
IA2*b*	nurturance to partner	18
IA2*c*	succorance from partner	17
IC6	war	13
IA1*g*	concern for child	10
IIA2*h*	dancing	10
IIA1*d*	reunion with parents	8
IA2*f*	departure from partner	7
IC5	aggression from environment	7
IIA2*d*	reunion with partner	7
IIB3	reminiscence, happy	7
IIA2*g*	anniversary	5
IA1*c*	nurturance to parent	3
IA1*f*	departure from parent	3
IA2*t*	restraint from partner	3
IIA1*b*	resignation to child	3

Card 11

IC5	aggression from environment	38
IC7	escape from perilous environment	18
IB9	fear, worry	15
IA3*e*	aggression to peer	13
IB3	curiosity	10
IB12	illness or death of central character	8
IA3*a*	pressure from peer	7
IA3*d*	aggression from peer	5
IB4	behavior disorder	5
IC3	generalized restriction	5
IC4	aggression to environment	5
IC1	economic pressure	3
IC6	war	3

Card 12F

IA1*l*	disappointment in parent	22
IA1*a*	parental pressure	20
IA1*b*	succorance from parent	10
IB14	reminiscence, sad	10
IB4	behavior disorder	8
IB7	guilt	8
IB31	regression	8
IB32	posing	8
IA1*c*	nurturance to parent	7
IB9	fear, worry	7
IA1*i*	death, illness of parent	5
IA1*k*	disappointment to parent	5
IA1*m*	filial obligation	5
IC1	economic pressure	5
IIB1	self-esteem	5
IA1*p*	marriage	3
IA1*q*	collusion with parent	3
IA2*m*	jealousy of partner	3
IA3*q*	devaluation of peer	3
IB2	inadequacy	3
IC5	aggression from environment	3
IIB3	reminiscence, happy	3

Card 13MF

IA2*j*	death, illness of female partner	78
IB7	guilt	42
IA2*e*	aggression from male partner	15
IA2*a*	pressure from partner	13
IC1	economic pressure	13
IB5	suicide	12
IA2*h*	illicit sex	10
IA2*i*	illicit sex with violence	7
IA2*o*	unfaithful male partner	7
IA2*u*	childbirth	7
IB2	inadequacy	7
IB4	behavior disorder	7
IB11	occupational concern	7
IC5	aggression from environment	7
IA1*j*	death, illness child	5
IB13	retribution	5
IA2*s*	unrequited male partner	3
IB6	moral struggle	3
IB8	drunkenness	3
IB16	religion	3
IB21	exhaustion	3

Card 14

IB1	aspiration	32
IIB7	ordinary activity	10
IB11	occupational concern	8
IIB2	tranquility	8
IC1	economic pressure	7
IIB3	reminiscence, happy	7
IA3*j*	belongingness	7
IA1*a*	parental pressure	5
IB4	behavior disorder	5
IB5	suicide	5
IB10	rumination	5
IB17	loneliness	5
IIC1	favorable environment	5
IA2*s*	unrequited male partner	3
IA2*t*	restraint from partner	3
IB2	inadequacy	3
IC3	generalized restriction	3
IC6	war	3

Card 15

IB15	intra-aggression	30
IB16	religion	30
IA2*j*	death, illness female partner	22
IA3*i*	death, illness friend	20
IB4	behavior disorder	18
IB17	loneliness	18
IB7	guilt	15
IB20	acquisition	15
IB14	reminiscence, sad	12
IB27	supernatural	12
IC4	aggression to environment	8
IA1*j*	death, illness child	8
IA3*e*	aggression to peer	5
IA3*k*	unappreciated by peer	5
IA3*p*	disregard for peer	5
IIB7	ordinary activity	5
IB29	illness or death of unspecified character	3
IA3*j*	belongingness	3
IA4*i*	death, illness sibling	3
IB9	fear, worry	3
IB10	rumination	3
IB12	illness, death central character	3

Card 16

IIA1*f*	contentment, parent	18
IB1	aspiration	12
IC5	aggression from environment	8
IIA2*c*	contentment with partner	8
IIA3*b*	congeniality with peer	7
IA1*a*	parental pressure	7
IA2*b*	nurturance to partner	7
IA1*i*	death or illness of parent	5
IB3	curiosity	5
IB10	rumination	5
IIA1*e*	fulfillment, parent	5
IB2	inadequacy	3
IIB2	tranquility	3
IIB7	ordinary activity	3

Card 17GF

IB5	suicide	32
IIB7	ordinary activity	32
IB3	curiosity	18
IC3	generalized restriction	12
IC1	economic pressure	10
IB1	aspiration	8
IC4	aggression to environment	8
IC5	aggression from environment	8
IA3*k*	unappreciated by peer	7
IA2*ss*	unrequited female partner	5
IB11	occupational concern	5
IB17	loneliness	5
IA1*a*	parental pressure	3
IA1*c*	nurturance to parent	3
IA2*s*	unrequited male partner	3
IA3*c*	nurturance to peer	3
IB4	behavior disorder	3
IIB5	occupational satisfaction	3

Card 18GF

IA1*b*	succorance from parent	20
IA1*a*	parental pressure	17
IA1*j*	death, illness of child	15
IA3*e*	aggression to peer	13
IA1*d*	aggression from parent	13
IB4	behavior disorder	13
IC5	aggression from environment	13
IA4*i*	death, illness of sibling	12
IC1	economic pressure	10
IA1*i*	death, illness of parent	8
IA4*e*	aggression to sibling	8
IA3*i*	death, illness of friend	7
IA4*c*	nurturance to sibling	7
IB9	fear, worry	7
IA1*c*	nurturance to parent	7
IA1*l*	disappointment in parent	7
IA1*g*	concern for child	5
IA3*b*	succorance from peer	5
IB7	guilt	5
IB12	physical illness, death of central character	5
IA1*f*	departure from parent	3
IA2*m*	jealousy of partner	3
IA2*oo*	unfaithful male partner	3
IA4*a*	pressure from sibling	3
IA4*g*	concern for sibling	3
IB3	curiosity	3
IB13	retribution	3
IB20	acquisition	3

Card 19

IC5	aggression from environment	70
IIA1*f*	contentment with parent	23
IB27	supernatural	20
IB9	fear, worry	8
IB4	behavior disorder	7
IB3	curiosity	5
IB17	loneliness	5
IC1	economic pressure	5
IIA3*b*	congeniality with peer	5
IIB2	tranquility	5
IA2*q*	pursuit of partner	3
IA3*b*	succorance from peer	3
IB2	inadequacy	3

Card 20

IB19	vacillation	38
IB17	loneliness	25
IC4	aggression to environment	17
IA3*b*	succorance from peer	15
IC5	aggression from environment	13
IB11	occupational concern	8
IB10	rumination	7
IA2*s*	unrequited male partner	7
IA3*e*	aggression to peer	7
IA3*k*	unappreciated by peer	7
IB3	curiosity	7
IB8	drunkenness	7
IB9	fear, worry	7

IA3*c*	nurturance to peer	5
IA3*j*	belongingness	5
IB2	inadequacy	5
IB14	reminiscence, sad	5
IC6	war	5
IIA2*c*	contentment	5
IIB7	ordinary activity	3
IC1	economic pressure	3
IB6	moral struggle	3

Table 5

Unusual Formal Characteristics Contributed by One or More Subjects for Each of 20 Cards

Code	Characteristics	Percentage frequency
Card 1		
III I	alternate themes	3
III G	autobiographical	2
III J	comments	2
Card 2		
III D	unreal	7
III I	alternate themes	3
III G	autobiographical	2
III J	comments	2
Card 3GF		
III I	alternate themes	5
III K	denial	3
III R	reification	3
III G	autobiographical	2
Card 4		
III D	unreal	10
III I	alternate themes	5
III K	denial	5
Card 5		
III I	alternate themes	3
III H	continuations	2
III K	denial	2
Card 6GF		
III I	alternate themes	5
III D	unreal	2
Card 7GF		
III I	alternate themes	5
III G	autobiographical	3
Card 8GF		
III D	unreal	3
III F	central character not in picture	3
III I	alternate themes	3
III R	reification	3
III C	descriptive	2
III H	continuations	2
Card 9		
III I	alternate themes	10
III G	autobiographical	2
III C	descriptive	2
III R	reification	2
Card 10		
III G	autobiographical	3
III J	comments	2
III R	reification	2
Card 11		
III C	descriptive	17
III D	unreal	13
III I	alternate themes	10
III E	fairy tale	8
III G	autobiographical	3
III R	reification	3
III A	symbolic	2
Card 12F		
III D	unreal	13
III I	alternate themes	8
III C	descriptive	7
III E	fairy tale	7
III K	denial	3
III G	autobiographical	2
III R	reification	2
Card 13MF		
III I	alternate themes	12
III G	autobiographical	2
III R	reification	2
Card 14		
III I	alternate themes	3
III H	continuations	2
III G	autobiographical	2
Card 15		
III D	unreal	10
III F	central character not in picture	7
III I	alternate themes	7
III C	descriptive	2
III J	comments	2
III K	denial	2
III R	reification	2
Card 16		
III G	autobiographical	22
III D	unreal	3
III C	descriptive	2
III F	central character not in picture	2
III H	continuations	2
III R	reifications	2
Card 17GF		
III K	denial	12
III I	alternate themes	7
III A	symbolic	3
III J	comments	3
III R	reification	3
III C	descriptive	2
III G	autobiographical	2
III L	rejection	2
Card 18GF		
III I	alternate themes	7
III K	denial	3
III G	autobiographical	2
III H	continuations	2

Card 19		
III D	unreal	17
III C	descriptive	12
III A	symbolic	8
III E	fairy tale	7
III J	comments	5
III F	central character not in picture	3
III G	autobiographical	3
III I	alternate themes	3
Card 20		
III I	alternate themes	8
III G	autobiographical	3
III R	reification	3
III A	symbolic	2
III K	denial	2

Table 6

Perceptual Distortions Contributed by One or More Subjects for Each of 20 Cards

Distortion	Percentage frequency
Card 1	
broken string	3
boy's eyes closed	2
Card 2	
*pregnancy noted	8
background is girl's imagination	3
Card 4MF	
woman in background seen as real, not picture	10
*woman in foreground seen as nurse	5
woman in foreground seen as Negress	3
*characters seen as brother and sister	2
Card 6GF	
*husband, wife (contemporaries)	68
*older husband	10
*father, daughter	7
*father or husband	3
*Sherlock Holmes	2
*boss	2
Card 7GF	
girl is holding a baby	8
uncertain if older woman is holding book	2
Card 8GF	
sexual misidentification	2
Card 9GF	
a reflection in the water	3
Card 10	
*elderly couple	25
*young couple	13
mother, son	7
two males	3
Card 12F	
sexual misidentification (figure in foreground)	3
Card 13MF	
*nudity of woman mentioned	8
*doctor, patient	7
father, daughter	3
Card 14	
individual sitting on window ledge	3
individual climbing in window	2
Card 15	
manacled hands	8
chairs instead of crosses	2
man in background	2
Card 17GF	
sexual misidentification	18
Negro character on bridge	5
don't mention character on bridge	2
seen as two different scenes	2
Card 18GF	
sexual misidentification	28
Card 20	
sexual misidentification	3
no head on man	2
no character seen	2

---

* These are not actually perceptual distortions. They are in this table in order to furnish normative data for unusual interpretations.

Table 7

Percentage of Subjects Contributing in a Protocol of 20 Stories a Given Number or More Stories of Each Emotional Tone Rating

Rating	Frequency in 20 stories														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
−2	98	90	73	33	15	5	3	2							
−1	100	100	100	97	92	88	75	60	40	22	15	5	2	2	2
0	100	93	87	80	57	42	33	25	15	8	3	2	2		
+1	100	73	52	37	17	10	7	3							
+2	28	2													

Table 8

Percentage of Subjects Contributing in a Protocol of 20 Stories a Given Number or More Stories of Each Outcome Rating

Rating	Frequency in 20 stories															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
−2	82	60	28	20	12	5	3	3	2	2	2	2				
−1	92	78	53	32	22	10	3									
0	82	67	37	17	10	8	7	3	2							
+1	97	90	83	73	62	45	38	27	15	12	7	5	5	3	3	2
+2	70	38	27	22	13	8	8	5	3							
?	40	13	7	2	2	2	2	2	2							
x	87	77	65	48	38	27	18	17	12	10	8	5	5	2		

Table 9

Percentages of Subjects Contributing in a Protocol of 20 Stories a Given Number or More of Every Theme Contributed at Least Twice by Any Subject

Code	Theme	Frequency in 20 stories					
		1	2	3	4	5	6
IA1*a*	parental pressure	87	62	48	20	5	2
IA1*b*	succorance from parent	77	13	3			
IA1*c*	nurturance to parent	25	3	2	2		
IA1*f*	departure from parent	18	3				
IA1*g*	concern for child	42	15	3			
IA1*i*	death or illness of parent	32	8	2			
IA1*j*	death or illness of child	45	8	2	2		
IA1*k*	disappointment to parent	23	5	3			
IA1*l*	disappointment in parent	32	5	2			
IA1*m*	filial obligation	25	3				
IA2*a*	pressure from partner	60	22	3			
IA2*b*	succorance from partner	42	12				
IA2*c*	nurturance to partner	45	12				
IA2*f*	departure from partner	38	8				
IA2*gg*	concern for female partner	12	2				
IA2*h*	illicit sex	35	5				
IA2*i*	illicit sex with violence	7	2				
IA2*j*	death or illness female partner	75	23				
IA2*jj*	death or illness male partner	23	2				
IA2*m*	competition for partner	30	5				
IA2*oo*	unfaithful to partner	30	3				
IA2*q*	pursuit of partner	18	2				
IA2*r*	seduction	27	3				
IA2*s*	unrequited male partner	18	2				
IA2ss	unrequited female partner	33	7				
IA2*t*	restraint	10	2				
IA2*u*	childbirth	10	2				
IA3*a*	pressure from peer	27	3				
IA3*b*	succorance from peer	30	5				
IA3*d*	aggression from peer	8	2				
IA3*e*	aggression to peer	32	8	2			
IA3*i*	death or illness friend	28	5				
IA3*j*	belongingness	47	10	3			
IA3*k*	unappreciated by peer	33	7	2			
IA3*n*	envy	8	2				
IA4*c*	nurturance to sibling	8	3				
IA4*f*	sibling rivalry	28	3	2			
IB1	aspiration	83	52	33	10		
IB2	inadequacy	47	13	3	2		
IB3	curiosity	80	45	15	2		
IB4	behavior disorder	62	12	7	5		
IB5	suicide	47	8	2	2		
IB6	moral struggle	17	2				
IB7	guilt-remorse	63	32	15	5	2	
IB9	fear-worry	60	37	8			
IB10	rumination	28	2				
IB11	occupational concern	42	37	12	7		
IB12	illness or death of central character	30	2	2			
IB14	reminiscence, sad	37	8				
IB15	intra-aggression	30	2				
IB16	religion	33	8	2			
IB17	loneliness	52	22	7	2		
IB19	vacillation	43	3	2			
IB20	acquisition	20	3				
IB21	exhaustion	18	2				
IB22	inconsequential sadness	17	2				
IB27	supernatural	30	3				
IC1	economic pressure	47	23	17	5	3	2
IC2	legal restriction	12	2				
IC3	general restriction	30	7				
IC4	aggression to environment	45	12				
IC5	aggression from environment	88	53	25	13	5	
IC6	war	28	8				
IC7	escape	50	15				

IIA1*a*	cooperation with parent	18	3	2		
IIA1*f*	contentment with parent	40	8	2		
IIA1*g*	ordinary familial activity	17	3			
IIA2*c*	contentment with partner	50	5	2		
IIA2*d*	reunion with partner	8	2			
IIA3*b*	congeniality with peers	20	2			
IIB1	self-esteem	13	5	2	2	
IIB2	tranquility	23	5			
IIB3	reminiscence, happy	40	10	2		
IIB7	ordinary activity	57	18	8	3	2
IIC1	favorable environment	23	3			

On only one out of 20 cards was there a difference in emotional tone rating between the two groups of subjects which was significant at or beyond the .05 level of confidence. Table 2 sets forth the card-by-card normative data for emotional tone. As regards outcome there were four significant card differences for the linear scale of values and no significant differences on any card for the (?) category. The outcome data are presented for each card in Table 3. In merely 10 instances were there significant differences in themes (total possible number of comparisons was 150×20=3000) and the card-by-card normative data appear in Table 4. There was a significant difference in unusual formal characteristics on only one card and there were no such differences on any single card for perceptual distortions. The normative data for these two variables are found in Tables 5 and 6 respectively.

When the data for all 20 cards were pooled, so that analysis was possible in terms of the frequency of individual subjects contributing a given number of a specific theme, emotional tone, distortion, formal characteristics, etc., in a total protocol of 20 stories, the two groups of subjects were still found to be quite similar. There were no differences in emotional tone of the stories, but the pregnant subjects did give significantly more happy outcomes than the college girl subjects (normative data in Tables 7 and 8 respectively). Out of a check list of 150 themes, on only one do these two groups differ—more college girls gave more themes of occupational concern than did the pregnant women (Table 9). There were no differences in unusual formal characteristics (Table 10) but more pregnant women produced more perceptual distortions, excluding sexual misidentifications, than did the college girls (Table 11).

From these data concerning significant differences it is apparent that both samples are really drawn from the same population insofar as TAT stories analyzed according to this methodology are concerned, since there were

Table 11

Percentages of Subjects Having a Given Number or More of Perceptual Distortions in a Protocol of 20 Stories

Type of distortion	Frequency in 20 stories 1	2	3	4	5
Sexual misidentifications	50	12	2		
Other distortions	57	32	5		
Total distortions	70	45	12	5	2

Table 10

Percentages of Subjects Contributing a Given Number or More Stories with Unusual Formal Characteristics in a Protocol of 20 Stories

		Frequency in 20 stories										
Code	Characteristics	1	2	3	4	5	6	7	8	9	10	11
IIIA	Symbolic	12	3									
IIIC	Descriptive	28	12	5	2	2						
IIID	Unreal	53	18	7	2							
IIIE	Fairy tale	22	2									
IIIF	Central character out	10	3									
IIIG	Autobiographical	22	5	3	2	2	2	2	2	2	2	
IIII	Alternate themes	45	25	17	10	2	2	2	2	2	2	2
IIIJ	Comments	10	2	2								
IIIK	Denial of theme	22	7	2								
IIIR	Reification	10	3	3	3	3	3	2				

hardly more than differences between the two groups than would be expected by chance in the number of implicit and direct comparisons made for each variable. Actually, then, it is reasonable to pool these data and offer norms based on a sample of 60 "normal" American women in this age range. By referring to these norms it will be possible to determine whether any given story contributed by a woman in response to a particular card is unusual or if there is a characteristic consistency throughout a series of 20 stories which is outside normal limits. These are the basic data—the frequency and unusualness of specific fantasy content — which must be determined before any interpretive significance can be placed on the thematic material obtained. How any TAT analyst will then use these data to interpret personality structure and content will depend on his own theoretic orientation and on the amount of behavioral data obtained from other sources which are available to him and in the light of which he can ascribe meaning to the fantasy material.

Received December 11, 1952.

References

1. Cohen, L. D. Stimulus properties of individual cards on the Thematic Apperception Test. Unpublished manuscript. Duke Univer., 1952.
2. Eron, L. D. Frequencies of themes and identifications in the stories of schizophrenic patients and non-hospitalized college students. *J. consult. Psychol.*, 1948, **12**, 387–395.
3. Eron, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, **64**, No. 9 (Whole No. 315).
4. Eron, L. D., & Ritter, Anne M. A comparison of two methods of administration of the Thematic Apperception Test. *J. consult. Psychol.*, 1951, **15**, 55-61.
5. Eron, L. D., Terry, Dorothy, & Callahan, R. The use of rating scales for emotional tone of TAT stories. *J consult. Psychol.*, 1950, **14**, 473-478.
6. Freedman, L. Z., Redlich, F. C., Eron, L. D., & Jackson, Edith. Training for childbirth: remembrance of labor. *Psychosom. Med.*, 1952, **14**, 439-452.
7. McNemar, Q. *Psychological statistics.* New York: Wiley, 1949.
8. Murray, H. A. *Manual for the Thematic Apperception Test.* Cambridge: Harvard Univer. Press, 1943.
9. Ritter, Anne M., & Eron, L. D. The use of the Thematic Apperception Test to differentiate normal from abnormal groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 147-158.
10. Rosenzweig, S. Apperceptive norms for the Thematic Apperception Test: I. The problem of norms in projective methods. *J. Pers.*, 1949, **17**, 475-482.
11. Terry, Dorothy. The use of a rating scale of level of response in TAT stories. *J. abnorm. soc. Psychol.*, 1952, **47**, 507-511.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Ambiguity of Pictures and Extent of Personality Factors in Fantasy Responses[1]

Douglas T. Kenny
University of British Columbia

and Sidney W. Bijou
University of Washington

This paper is the second in a series reporting on the relationship between the ambiguity qualities of picture stimuli and extent of personality factors in the elicited fantasy. The first [2] described the method used for establishing ambiguity values for Thematic Apperception Test pictures. It is the object of this paper to describe the findings on the relationship between stimulus ambiguity of Thematic Apperception Test pictures and extent of personality factors evaluated from fantasy productions.

We have pointed out [2] that a common assumption in social and clinical psychology is that the more ambiguous the stimulus the greater the apperceptive distortion, hence the greater the amount of personality factors revealed. The problem of how the stimulus structure affects the extent of personality factors "determining" a response is central to any behavior theory attempting to interrelate stimulus ambiguity on one hand, and perception, apperception, and motivational states on the other. Most experiments in this area attempt to show how "inner" factors (intervening variables) play a role in perception and apperception rather than how configurational ambiguity is related to the extent of personality factors "injected" into the stimulus situation.

The hypothesis under consideration has provided part of the rationale for the construction of the TAT and other projective techniques. And because of the obvious pertinence of this hypothesis to clinical, social, and personality theory and to the construction of clinical instruments as well, there is a continual need for research which aims to evaluate its validity.

Method

Ascertaining ambiguity level. The first paper [2] in this series may be referred to for a discussion of alternative methods for measuring the ambiguity qualities of stimulus material for obtaining apperceptive responses. We will recapitulate briefly our approach to the problem. Data on the ambiguity of 21 TAT cards (the general and male series) were obtained by having 51 judges rank order the cards according to our definition of ambiguity. We defined ambiguity level in terms of the estimated number of interpretations a group of judges thought would apply to each TAT card. An ambiguity value for each card was therefore expressed in terms of its median rank-order value. The multi-judge reliability of the ranking was adequate ($\rho = .80$).

By this method we obtained three sets of five TAT pictures each, varying in three degrees of ambiguity. Cards 12BG, 1, 2, 9BM, and 17 BM had low median rank values and are therefore classified as least ambiguous. Cards 4, 7 BM, 8BM, 12M, and 13 MF were given intermediate rank values and make up the intermediate ambiguity group. And cards 5, 15, 18 BM, 11, and 19, which had the high rank values, comprise the most ambiguous set.

Obtaining fantasy responses. The 15 cards were administered individually to 18 male college students from introductory psychology classes at the University of Washington. The subjects were assigned at random to three comparably trained examiners, two males and one female. Fantasy reactions to the cards were obtained according to Murray's [8] procedures with certain necessary modifications. In order to present comparable test situations, test directions were standardized. The experimental design required the subject to relate a story to each card of a set on three successive days, the cards being administered in random order between and within sessions. Stories were recorded on a dictaphone machine. Each story was then typed on a single sheet of paper and assigned a code number. From our 18 subjects we obtained 270 stories, 90 for each of the three sets of cards. Before being presented to the judges, the typewritten protocols were shuffled to eliminate the influence of fixed sequential effects.

---

[1] The authors wish to express their appreciation to Dr. Robert Quinn and Dr. Ira Steisel for so conscientiously performing the arduous task of judging the protocols.

Measuring extent of personality factors. To quantify the significance of personality factors in the TAT protocols, we used a modification of Stephenson's Q-sort technique [9]. The material consisting of the 270 stories from 18 subjects was given to the judges in two groups of 135 each. The judges, two well-qualified clinical psychologists working in different mental hygiene clinics, were asked to sort the stories in terms of the amount of personality factors revealed by placing each in one of nine categories, with the following normal frequency distribution:

Score	1	2	3	4	5	6	7	8	9
Frequency	7	11	16	20	27	20	16	11	7

Stories judged to have little personality material were given a value of 1, stories with a great deal of personality material, a value of 9.

Directions to the two judges were as follows:

> You are requested to rate 135 TAT stories on the basis of the amount of personality structure and dynamics (motives and motivational structure, attitudes, defenses, conflicts, self concepts, values, goals, etc.) revealed about the subject.
>
> Please read each story carefully and then place it in one of three piles. Place in Pile 1 all those stories you believe show few or no personality factors. Place in Pile 2 all those stories you believe show some personality features (moderately rich in personality content). Place in Pile 3 all those stories you believe show considerable or extensive personality information.
>
> Review the stories in Pile 1 (low in personality content) and select the 7 stories you believe to have the least amount of personality information. Place these in Group 1. Review the stories in Pile 3 (high in personality content) and select the 7 stories you believe to have the greatest amount of information. Place these in Group 9. Review the stories remaining in Pile 1 and select the 11 stories you believe to have a little more information than those placed in Group 1. Place these in Group 2. Review the stories remaining in Pile 3 and select the 11 stories you believe to have considerable information but not so extensive as those placed in Group 9. Place these in Group 8. Review the stories remaining in Pile 1 and place the next 16 stories in Group 3. These would be more informative than those you placed in Group 2. Review the stories remaining in Pile 3 and place the next 16 stories in Group 7. These should be relatively rich in content but not quite so informative as those placed in Group 8.
>
> Handle the middle group the same way. First place the 20 stories with the least amount of personality information in Group 4. Then place the 20 stories with the greatest amount of information (within this group) in Group 6. The remaining 27 stories will constitute Group 5.

Results

Differences in amount of fantasy productions. The first question likely to arise is: Is there a relationship between length of story and ambiguity level of a card? A word count was made of the 90 stories obtained at each of the three levels of ambiguity. The mean number of words composing the stories from the least ambiguous cards was 980.2, for the med-

Table 1

Differences in Story Length Related to Test Administrator

Test administrator	Mean number of words	Mean difference	SE of mean difference	t
L vs. J	1272.9 644.3	628.6	188.76	3.33**
L vs. W	1272.9 796.2	476.7	190.78	2.50*
W vs. J	796.2 644.3	151.9	146.37	1.04

** Significant at 1 per cent level.
* Significant at 5 per cent level.

Table 2

Analysis of Variance of Values from a Q-Sort of Stories Based on Personality Content

Source of variation	Sum of squares		*df*		Mean square	*F*
Examiners	1036.50		2		518.75	4.76*
Subjects in same group	1636.19		15		109.08	
Total between subjects		2672.69		17		
Raters	.00		1		.00	....
Ambiguity level	877.17		2		438.59	12.19**
Examiners × Judges	7.39		2		3.70	....
Examiners × Ambiguity	138.66		4		34.67	....
Ambiguity × Judges	45.50		2		22.75	....
Examiners × Ambiguity × Judges	62.39		4		15.60	....
Pooled subjects × Ambiguity	2692.20		75		35.90	....
Total within subjects		3823.31		90		
Total		6496.00		107		

** Significant at 1 per cent level.
* Significant at 5 per cent level.

ium ambiguous cards 947.7, and for the most ambiguous 785.5. None of the differences between the means was statistically significant.

Our research design produced some evidence on the question of whether the length of the stories varied as a function of the behavior (personality) of the test administrators. It will be recalled that there were three examiners, one female, two males. Word counts were made of the stories obtained by each examiner. Tests of the reliability of the differences between pairs of means indicated that one of the male examiners (L) obtained reliably longer stories than the other two examiners. These findings are presented in Table 1.

Differences in content "richness" of fantasy productions. The next question, and from the point of view of this paper, the most important, is: Are there differences in personality content in the protocols as a function of the ambiguity level of the cards? Before presenting the findings bearing on this question, a brief comment on judge reliability may be helpful.

The correlation of the scores of the two judges is .68, which is statistically significant at the 1 per cent level of confidence. While the correlation does provide a measure of agreement between the raters, it does not reflect accurately the reliability of the pooled ratings used in the analysis. Applying the Spearman-Brown prophecy formula, the correlation of .68, the estimate of the reliability of the combined judgments becomes .81.

An analysis of variance was made to determine not only the effect of ambiguity level on amount of personality material revealed but also the effect of examiner and judge differences. The results of this analysis are shown in Table 2.

Two factorial effects seem to be significant: level of ambiguity of the cards and examiner differences. However, neither the judge effects

Table 3

Differences in Content Scale Value Related to Test Administrator

Test administrator	Mean	Mean difference	*SE* of mean difference	*t*
L	5.82			
vs.		1.50	.66	2.27*
J	4.32			
L	5.82			
vs.		.91	.67	1.36
W	4.91			
W	4.91			
vs.		.59	.55	1.07
J	4.32			

* Significant at 5 per cent level.

nor any of the interactions approaches significance. To determine which examiners and which sets of cards showed differences a further analysis was made by means of *t* tests. Since the performance of the two judges did not show significant differences, their scores were pooled. Calculations of *t* tests are, therefore, in terms of the combined judgments.

Turning first to the analysis of the examiner differences we find that there was no significant difference between the two male examiners and that one of the male examiners obtained stories similar to those obtained by the female examiner. The second male examiner's subjects related stories containing higher personality values than those of the female examiner. These findings are summarized in Table 3.

Finally, examination of the personality-score differences as related to picture ambiguity, given in Table 4, shows that the cards in the medium ambiguous set were significantly higher than those in the least and most ambiguous sets. There is no statistical difference between the mean values of the least and of the most ambiguous groups of pictures.

Table 4

Differences in Content Scale Values Related to Ambiguity Level of TAT Cards

Ambiguity level	Mean scale value	Mean difference	*SE* of mean difference	*t*
Least	4.57			
vs.		1.24	.54	2.30*
Medium	5.81			
Least	4.57			
vs.		.05	.49	.10
Most	4.62			
Medium	5.81			
vs.		1.19	.39	3.05**
Most	4.62			

** Significant at the 1 per cent level.
* Significant at the 5 per cent level.

Our material yielded one other interesting bit of information. It will be recalled that Murray [9] in constructing the TAT divided the cards into two sets of ten. Cards numbered 10 and below were believed to be more structured than those numbered 11 and above. Of the 15 cards used in this study, 7 are from the first set of 10, and 8 are from the second set of 10. To determine whether there was any difference between Murray's estimate of ambiguity and extent of personality factors revealed, the data were arranged so that a mean pooled rater judgment was obtained for the 7 cards belonging to the first set of 10 and, similarly, a mean pooled rater judgment for the 8 cards belonging to the second set of 10. The mean "personality richness value" for the cards from the first 10 cards was 5.22 and that for the cards in the second 10 cards was 4.82. The *t* value equaled .93, which was not statistically significant. Hence, when the cards are arranged according to Murray's classification, we do not find his most ambiguous cards eliciting more significant personality factors than his least ambiguous cards. The possibility that the second series of pictures often elicits more meaningful material for personality analysis may be primarily a function of the change in test directions and the influence of card sequence rather than the more "dramatic" qualities of the second series of pictures.

Discussion

This study presents interesting findings not only on the problem of stimulus ambiguity and fantasy productions but also on examiner behavior and fantasy productions. One of the male examiners obtained protocols that were longer than those secured by the other two examiners. The stories given to the same examiner were also richer in personality content than those told to the female examiner. It is apparent that interpersonal stimulations between examiner and subject had differential effects despite efforts to standardize the instructions and the conduct of the examiners. For example, the examiners were required to read the test instructions, refrain from making spontaneous comments during the tests, avoid conversation before and after the administration of the tests, record the stories on the dictaphone, and the like.

Clinical psychologists have long suspected that what is obtained from projective technique is in part a function of the personality of the administrator. Recently several investigators have demonstrated this to be so, namely, Baughman [1], Gibby, *et al.* [5], and Lord [6] working with the Rorschach technique, and Milam [7] with the Thematic Apperception Test. The implication from such studies is obvious. Studies in this area need to be designed to control or at least to evaluate influences from this source. Unfortunately this is not a simple task. It is conceivable that such stimuli could be so subtle that they would be "blind spots" not only to the examiners but also to the investigators.

Our finding on stimulus ambiguity and extent of personality information revealed in fantasy responses when taken at face value is not consistent with the hypothesis that the two terms are related in a monotonic fashion. The five most ambiguous cards produced stories that were no richer in personality content than those from the five least ambiguous cards. Moreover, cards with median ambiguous values yielded stories with more personality information than those of the other two sets. It might be argued that this finding should not be taken as evidence refuting the monotonic hypothesis because of the way the terms were defined and measured. To this position we offer two comments. First, it is axiomatic that any obtained relationship is a function of the methods of measurements involved. If we had measured either one or both of our terms differently, we might have obtained different results.

The second point relating to this argument is the fact that the hypothesis does not include definitions of terms, operational or otherwise. Each investigator is therefore forced to define the terms in accordance with his own judgment. What is being evaluated here, then, is the hypothesis with concepts measured as described.

Let us now turn to a consideration of the findings on stimulus ambiguity and personality content. Our approach is in terms of drive strength and stimulus generalization [4]. In regard to drive strength, we may infer that the 18 male college students serving as subjects (*a*) had a medium-to-weak drive level as compared to a disturbed or deprived group; and (*b*) revealed the dominant drive as varying over a wide range. For example, the dominant drive for one individual may have been achievement, for another, sex, for still another, aggression, and so on. If this were so, we can then assume that the tendencies for perceptual responses instigated by the drives of the group were not very strong and were heterogeneous in kind.

Now let us consider the stimulus properties of the 15 cards. We could assume that the card stimuli varied in at least two dimensions: (*a*) the number of cue-constellations available for selective reactions, and (*b*) the general definiteness level of the cue-constellations. According to this formulation cards in set A would have a low frequency of cue-constellations. In addition most of these cue-patterns would be sufficiently definite to evoke similar perceptual responses from members of the group. On the other hand, the most ambiguous cards, set C, would have a high frequency of cue-constellations and most of them would vary in definiteness from cue-patterns that yield a high degree of perceptual response agreement. The cards in set B would, of course, have intermediate stimulus properties between those in set A and set C.

According to this formulation, cues in set A should produce response tendencies which compete favorably with those from the dominant drives. These stories should be "card dominated." They should consist of the symbolizing (and other) processes associated with the interpreted meanings of the cards. The hypothesis of how the story evolves is somewhat like this: In compliance with test instructions the individual selects and makes a perceptual reaction to an aspect of the card. He then relates organized verbal material pertinent to the need-satisfaction sequences, the conflicts, frustrations, gratifications, and so on. Stories based primarily on card stimuli would not be expected to contain much about the individual's dominant drive and the related personality concepts. By the same token, cards in set B should evoke perceptual response tendencies about equal in strength to those from the drives. In these situations sometimes the picture stimuli dominate, at other times the drive stimuli govern. In any case, many "perceptual errors" could be made. Numerous stories would therefore be based on the dominant drives. Such protocols should yield more information about personality than those from the pictures in set A. This is in agreement with the findings. Finally the cues in the pictures of set C should elicit weak response tendencies which would easily be overridden by response tendencies of the major drives. Stories from this set of pictures should be "drive dominated" and should reveal most about personality factors. This expectation is not in agreement with the findings.

What are some of the ways of accounting for the discrepancy between hypothesis and fact? (*a*) The hypothesis needs to be modified to account for another variable. It is possible, for example, that an uncontrolled influence was operating to lower fantasy behavior in response to cards in set C, which were the most ambiguous pictures. Some investigators [3] have suggested that as situations become increasingly vague, anxiety or stress is increased. Such a situational condition may have a general inhibitory effect on the performance to the least structured cards. (*b*) The cue-patterns in the cards do not vary as assumed—increasing ambiguity is correlated with an increasing number of cue-patterns and a decreasing clarity of cue-patterns. It is conceivable that there was a low relationship between either one or both of these dimensions and the judges' estimates of number of possible interpretations [2]. (*c*) The cue-patterns do vary as assumed, but there were qualitative differences among them. For instance, cards in set C may have had a greater number of cue-patterns than those in set B, but the latter had more relating to drives and situations of an interpersonal nature. In line with this comment, it should be pointed out that there were more human figures in set B than in set C. (*d*) Finally, the judges, in accordance with their training and experience as clinical psychologists, were inclined to give greater weight to protocol contents having to do with drives and dynamics associated with behavior pathology such as sex and aggression, and to underevaluate those pertaining to the milder and more "normal" processes such as affiliation and achievement. Presumably material concerning either set of terms would be revealing about personality, but because of the judges' orientation the pathological could have rated higher than the normal. Such a bias could have had a differential effect considering the content of the cards from set to set.

Summary

This paper describes an experiment testing the assumption of a direct relationship between stimulus ambiguity and extent of personality factors revealed in evoked fantasy. On the basis of findings in an earlier study, three sets of TAT cards varying in degree of ambiguity were selected and given to 18 young adult male subjects. Two well-qualified clinical psychologists, utilizing a Q-sort technique, judged the significance of the personality factors injected into the fantasy stories.

The results from an analysis of variance of the Q-sort scores are not consistent with the hypothesis under investigation. With increasing stimulus ambiguity there is an initial increase in the extent of personality factors revealed in the fantasy, and then a decrease. In addition, it was shown that the behavior of the examiner plays a role in fantasy productions. It was also evident that with instructions held constant, cards from Murray's second series did not reveal more information about personality than cards from his first series.

The hypothesis and the findings are analyzed in terms of drive strength and stimulus generalization.

Received December 1, 1952.

References

1. Baughman, E. E. Rorschach scores as a function of examiner difference. *J. proj. Tech.*, 1951, **15**, 243–249.
2. Bijou, S. W., & Kenny, D. T. The ambiguity values of TAT cards. *J. consult. Psychol.*, 1951, **15**, 203–209.
3. Cameron, N., & Magaret, Ann. *Behavior pathology.* Boston: Houghton Mifflin, 1951.
4. Dollard, J., & Miller, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
5. Gibby, R. G., Miller, D. R., & Walker, E. L. Examiner variance in the Rorschach protocols of neuropsychiatric patients. *Amer. Psychologist*, 1952, **7**, 337–338. (Abstract)
6. Lord, Edith. Experimentally induced variations in Rorschach performance. *Psychol. Monogr.*, 1950, **64**, No. 10 (Whole No. 316).
7. Milam, J. R. Examiner influences on thematic apperception stories. Unpublished master's dissertation, Univer. of Washington, 1952.
8. Murray, H. A. *Manual for the Thematic Apperception Test.* Cambridge: Harvard Univer. Printing Office, 1943.
9. Stephenson, W. Methodological considerations of Jung's typology. *J. ment. Sci.*, 1939, **85**, 185–205.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Measuring Achievement Motivation with "Picture Interpretations"

Henry H. Morgan
Psychological Corporation

The measurement of achievement motivation is the subject of recent research which has been reported by McClelland and others [2, 3, 4, 6]. This research employed a short version of the Thematic Apperception Test (TAT) in which stories written in response to picture slides were scored for Murray's need for Achievement, or *n Ach* [7, pp. 144, 164]. In the initial studies by McClelland *et al.* [4], groups of subjects took the test under experimentally induced motivational states, and a scoring system was evolved which made use of differences found between those taking the test under "ego-involving" conditions and those taking it under "neutral" conditions.

The study about to be described is an outgrowth of the research mentioned above and is part of an Office of Naval Research project undertaken to develop a standardization test of achievement motivation.[1] One purpose of the study was to determine the relationship between scores for achievement motivation, or *n Ach*, and a criterion of achievement which, in this case, was school grades. Another purpose was to test the stability of *n Ach* scores over a period of time.

Procedure

The sample. The subjects tested were boys in the junior class of high schools in a New England industrial city. Five groups of boys were used, ranging in number from 39 to 62 subjects. Groups 1, 2, and 3 were from a large academic high school, while Groups 4 and 5 were from a vocational high school where shop courses contributed to the grade criterion.

Students had taken the Otis Quick-Scoring Mental Ability Test (Beta test, Form Cm) during their freshman year in high school, and Otis IQ's were available for a majority of the subjects. Each academic high school group had an average Otis IQ of 109, while each vocational high school group had an average IQ of 99 on the Otis.

The measuring instrument. The achievement motivation test under development has been designated "Picture Interpretations," or P-I, to distinguish it from Murray's TAT. From a pool of pictures, some of which had been pretested in other high school populations,[2] four forms of P-I were constructed. These forms, incorporating 36 different pictures, were designated E, F, G, and H. Forms E, F, and G, which were made up of 12 pictures each, used plain unstructured response sheets (8½ in. × 14 in.) for the recording of the stories by the subjects. Form H, which contained the first eight pictures of Form G, had the same over-all time limit of 30 minutes for the test but used slightly different response sheets on which were printed a series of "story structuring" questions used in previous *n Ach* studies [4]. In all groups, subjects viewed each slide projection for a period of ten seconds. In the administration of Forms E, F, and G two and one-half minutes were allowed after each slide presentation for story writing, but in the administration of Form H three and three-quarters minutes were allowed for the writing of each story.

Subjects were told that the testing program was sponsored by the Office of Naval Research, that the test results would be used for research purposes only, and that students' responses would be kept anonymous. Each response booklet contained mimeographed instructions for the test, and these instructions were read aloud to the subjects before each test administration. The instructions for Forms E, F, and G are given below. Those for Form H were slightly different owing to the altered time limits.

---

[1] Research Project NR 172-116, Contract No. N onr 559(03), 1951-52, was under the direction of D. C. McClelland, Wesleyan University. Contract personnel included R. A. Clark, H. H. Morgan, R. W. Freeman, and R. W. Lull.

[2] Forms A, B, C, and D of P-I yielded correlations of *n Ach* scores and grades that ranged from .22 to .38. With intelligence partialed out, correlations of *n Ach* scores and grades ranged from .13 to .38.

Instructions

You are going to see a series of pictures, and your task is to tell a story that is suggested to you by each picture. Try to imagine what is going on in each picture. Then tell what the situation is, what led up to the situation, what the people are thinking and feeling, and what they will do.

In other words, write as complete a story as you can—a story with a plot and characters.

You will have ten seconds to look at a picture and then two and one-half minutes to write your story about it. Write your first impressions and work rapidly. I will keep time and tell you when it is time to finish your story and to get ready for the next picture.

There are no right or wrong stories or kinds of stories, so you may feel free to write whatever story is suggested to you when you look at a picture. Spelling, punctuation, and grammar are not important. What is important is to write out as fully and as quickly as possible the story that comes into your mind as you imagine what is going on in each picture.

Notice that there is a half page for writing each story and that a picture number appears on each half page. If you need more space for writing any story, use the reverse side of the paper.

In the administration of all forms a warning, "Try to finish your story in 30 seconds," was given one-half minute before expiration of the allotted writing time.

Data collection and analysis. Forms E, F, and G of P-I were administered to Groups 1, 2, and 3, respectively, in March, 1952, and each group was given either Form F or Form E as an alternate form five weeks later in April. Groups 4 and 5 were tested with Forms G and H, respectively, in February, 1952.

The protocols were analyzed for *n Ach* by two scorers using P-I scoring system "D",[3] a modification of system "C" [4]. A reliability coefficient of .89 was obtained for a sample of protocols analyzed by each of the scorers.[4]

Results

Product moment correlation coefficients between *n Ach* scores, grade averages, and intelligence test scores are given in Tables 1 through 5. Tables 1, 2, and 3 present the results for the academic high school groups, while the findings for the vocational high school groups are given in Tables 4 and 5.

Correlations in Tables 1 through 5 which are .39 or greater are significant beyond the .01 level of confidence using a CR technique [5, p. 122]. Correlations of .29 to .36 are significant beyond the .05 level. Correlations that are .27 or lower do not reach the .05 level of confidence. The partial correlations, which are between *n Ach* scores and school grades with Otis IQ partialed out [5, p. 141], give an estimate of the relationship between *n Ach* and the achievement criterion if intelligence is held constant. The subscripts of the partial *r*'s refer to the variables listed in the correspondingly numbered columns or rows in any given table.

Examination of Tables 1 through 5 reveals that, in general, low to moderate relationships

Table 1

Correlations between IQ, School Grades, and *n Ach* Scores among Junior Boys in an Academic High School, Sample No. 1 ($N = 53$ to 62)

(Form F administered 5 weeks after Form E)

Variable	1 Otis IQ	2 Sophomore Grade Average	3 Junior Grade Average	4 *n Ach* (Form E)	5 *n Ach* (Form F)
1. Otis IQ	----	.42	.45	.03	.09
2. Sophomore Grade Average	.42	----	.81	.16	.19
3. Junior Grade Average	.45	.81	----	.25	.29
4. *n Ach* (Form E)	.03	.16	.25	----	.56
5. *n Ach* (Form F)	.09	.19	.29	.56	----

Partial Correlations: $r_{42.1} = .15$; $r_{43.1} = .26$; $r_{52.1} = .17$; $r_{53.1} = .28$.

---

[3] McClelland, D. C., Clark, R. A., and Morgan, H. H., Revisions in the system of scoring imaginative stories for need: Achievement (*n Ach*); scoring system D (April, 1952). Unpublished paper.

[4] Protocols for Groups 1, 2, and 3 were scored by the author, those for Groups 4 and 5 by R. W. Freeman. Interscorer reliability on 20 protocols from Group 3, Form G, was .89.

Table 2

Correlations between IQ, School Grades, and *n Ach* Scores among Junior Boys in an Academic High School, Sample No. 2 ($N = 39$ to 53)

(Form E administered 5 weeks after Form F)

Variable	1 Otis IQ	2 Sophomore Grade Average	3 Junior Grade Average	4 *n Ach* (Form F)	5 *n Ach* (Form E)
1. Otis IQ	----	.16	.56	.24	.73
2. Sophomore Grade Average	.16	----	.65	.45	.43
3. Junior Grade Average	.56	.65	----	.47	.29
4. *n Ach* (Form F)	.24	.45	.47	----	.56
5. *n Ach* (Form E)	.73	.43	.29	.56	----

Partial Correlations: $r_{42.1} = .43$; $r_{43.1} = .42$; $r_{52.1} = .46$; $r_{53.1} = -.21$.

Table 3

Correlations between IQ, School Grades, and *n Ach* Scores among Junior Boys in an Academic High School, Sample No. 3 ($N = 51$ to 61)

(Form E administered 5 weeks after Form G)

Variable	1 Otis IQ	2 Sophomore Grade Average	3 Junior Grade Average	4 *n Ach* (Form G)	5 *n Ach* (Form E)
1. Otis IQ	----	.35	.50	.22	.13
2. Sophomore Grade Average	.35	----	.70	.41	.28
3. Junior Grade Average	.50	.70	----	.39	.24
4. *n Ach* (Form G)	.22	.41	.39	----	.64
5. *n Ach* (Form E)	.13	.28	.24	.64	----

Partial Correlations: $r_{42.1} = .36$; $r_{43.1} = .33$; $r_{52.1} = .25$; $r_{53.1} = .20$.

exist between *n Ach* scores and grades. There are also low, moderate, and high relationships between *n Ach* scores and intelligence test scores as well as substantial relationships between intelligence test scores and school grades. With intelligence partialed out, relationships of *n Ach* scores and school grades tend to be somewhat reduced.

Table 4

Correlations between IQ, School Grades, and *n Ach* Scores among Junior Boys in a Vocational High School, Sample No. 4 ($N = 45$ to 50)

Variable	1 Otis IQ	2 Sophomore Grade Average	3 Junior Grade Average	4 *n Ach* (Form G)
1. Otis IQ	----	.27	.32	.26
2. Sophomore Grade Average	.27	----	.70	.04
3. Junior Grade Average	.32	.70	----	.15
4. *n Ach* (Form G)	.26	.04	.15	----

Partial Correlations: $r_{42.1} = -.03$; $r_{43.1} = .07$.

Table 5

Correlations between IQ, School Grades, and *n Ach* Scores among Junior Boys in a Vocational High School, Sample No. 5 ($N = 40$ to 48)

Variable	1 Otis IQ	2 Sophomore Grade Average	3 Junior Grade Average	4 *n Ach* (Form H)
1. Otis IQ	----	.55	.46	.36
2. Sophomore Grade Average	.55	----	.67	.31
3. Junior Grade Average	.46	.67	----	.27
4. *n Ach* (Form H)	.36	.31	.27	----

Partial Correlations: $r_{42.1} = .14$; $r_{43.1} = .09$.

Tables 1, 2, and 3 give the correlations between alternate forms of P-I administered five weeks apart. These correlations of .56, .56, and .64 indicate that *n Ach* scores are fairly stable over a short period of time.

Discussion

The correlations between *n Ach* scores and grades vary from one group to another, and these variations may be due to several causes. In the first place, sample sizes are limited, ranging from 39 to 61 subjects, and therefore random fluctuations of considerable size may be expected among the coefficients computed in the various groups or samples. Some fluctuations fall well within *SEr* which, in most cases, approximates .14. In addition to unreliability due to the limited size of the samples, a certain amount of unreliability is probably inherent in the grade criterion, whose reliability can be estimated by the correlations between sophomore and junior grades which range from .65 to .81. Furthermore, the reliability coefficients of .56, .56, and 64 based on alternate forms of the achievement motivation test show that the P-I technique itself is another source of unreliability. One more factor accounting for some of the variation in corrrelations from one sample to another may be the differential effectiveness of the various test forms used in measuring *n Ach*.[5]

In this study, the *n Ach* scores derived from alternate P-I forms appear, for projective tests, to be moderately reliable. However, the relationships of *n Ach* scores to grades and to intelligence appear less stable, and predictions on the basis of individual *n Ach* scores may not be made with confidence at the present stage of test development. Since achievement motivation, as measured by *n Ach* score, might be expected to vary somewhat from one time to another in the same individual, future P-I test reliabilities may not rise very much above those reported in this study.[6] These reliabilities demonstrate that *n Ach* score may be considered a measurement of an enduring personality trait just as the initial experimental studies of McClelland *et al.* [4] showed that *n Ach* score reflects temporarily induced motivational states.

A number of writers [1, 8] believe that, in work with projective techniques, standard test-retest or split-half reliabilities are of less importance than interscorer reliabilities which test the stability of interpretation from one investigator to another. The interscorer reliability of .89 reported in this study, as well as higher figures reported by McClelland [4], show that the *n Ach* scoring system is a highly consistent one.

The results of this study and other P-I studies show that the test, which began as a projective technique for the experimental study of motivation, holds some promise as a predictive instrument for forecasting achievement.

Summary

A story-telling test known as Picture Interpretations, or P-I, was administered to five groups of junior boys in academic and vocational high schools. Stories written in response to a series of pictures were scored for achievement motivation or need for Achievement (*n Ach*).

Low to moderate relationships were found between *n Ach* scores and school grades and also between *n Ach* scores and intelligence. The relationship between *n Ach* scores and school grades tended to be somewhat less when intelligence was held constant by use of the partial correlation technique.

Alternate forms of P-I given five weeks apart yielded reliability coefficients of .56, .56, and .64 indicating *n Ach* scores to be fairly stable over brief periods of time.

Received November 10, 1952.

References

1. Harrison, R., & Rotter, J. B. A note on the reliability of the TAT. *J. abnorm. soc. Psychol.*, 1945, **40**, 97–99.
2. Lowell, E. L. The effect of need for achievement on learning and speed of performance. *J. Psychol.*, 1952, **33**, 31–40.
3. McClelland, D. C. *Personality*. New York: William Sloan, 1951.
4. McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953, in press.
5. McNemar, Q. *Psychological statistics*. New York: Wiley, 1949.
6. Morgan, H. H. A psychometric comparison of achieving and nonachieving college students of high ability. *J. consult. Psychol.*, 1952, **16**, 292–298.
7. Murray, H. A. *Explorations in personality*. New York: Oxford Univer. Press, 1938.
8. Rosenzweig, S. Idiodynamics in personality theory with special reference to projective methods. *Psychol. Rev.*, 1951, **58**, 213–223.

---

[5] Further analysis of the data has been undertaken by Dr. Henry N. Ricciuti, Educational Testing Service, to determine which pictures and which components of the *n Ach* scoring system are contributing consistently to the relationship between *n Ach* scores and achievement criteria.

[6] Split-half reliability coefficients (Spearman-Brown) on Forms A, B, C, and D ranged from .36 to .77.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Social and General Intelligence and Performance on the Rosenzweig Picture-Frustration Study

Lawrence Karlin
New York University

and Milton M. Schwartz
Rutgers University

Inspection of the various social situations depicted by the Rosenzweig Picture-Frustration (P-F) test suggested the possibility that social intelligence, aside from other determinants such as motivation, needs, etc., may play a role in determining the type of response that is made to these situations. The purpose of the present study is to explore this possibility by analyzing the pattern of interrelationships obtained between the scoring categories of the Rosenzweig P-F test and the subtests of a well-known social intelligence test, in this case the George Washington University Series Social Intelligence Test.

To some extent at least, responses to the various situations on the P-F test might require a certain degree of inventiveness and social insight and thus one might expect a positive relationship between the occurrence of responses indicating these characteristics on the P-F test and the subtest purportedly measuring the same characteristics on the Social Intelligence test. Furthermore, indications of behavioral difficulties obtained on the P-F test might correlate with lack of social skills as measured by the Social Intelligence test. Thus, the comparison of the pattern of performance obtained on each might enhance the understanding of the projective material furnished by the P-F test.

To complement this comparison it was decided to have a conventional test of general intelligence, the Otis Self-Administering Tests of Mental Ability. It was felt in such an exploratory study that the general hypothesis centering around the influence of social skills which might make use of a person's inventiveness and insight would also be reflected in significant relationship with a test of general intelligence.

Subjects and Materials

The three tests were administered to 45 male students in introductory psychology classes at Rutgers University at two-week intervals in the following order: Otis Test of Mental Ability, Rosenzweig P-F test, and the Social Intelligence test. The following is a brief description of the tests employed:

1. Rosenzweig Picture-Frustration Study (Revised Form) [3]. This is a semiprojective personality test consisting of 24 picture situations portraying in cartoon-like fashion typically encountered "frustrating" experiences. In each picture there is depicted conversation between the person producing the frustration and the one presumably being frustrated. The subject is required to write, in the space provided, his first response to the frustrating statement expressed. Scoring is in terms of the standard procedure of Rosenzweig which consists of seven scoring categories; three involve Direction of Aggression (Outward, Inward, and Nonaggression) and four Types of Reaction, (Obstacle-dominance, Ego-defense, Need-persistence, and Denial).

2. The George Washington University Series Social Intelligence Test of Moss, *et al.* [1]. This test consists of five subtests, namely, Judgment in Social Situations, Recognition of the Mental State of the Speaker, Memory for Names and Faces, Observation of Human Behavior, Sense of Humor. Each test may be scored separately and the scores may be combined into a total score which represents the sum of the subtests. These will be described under results wherever necessary to elucidate the meaning of the findings of this study. Suffice it to say at this point that the literal meanings of each of the five titles provides a good idea of the content of each.

3. The Otis Self-Administering Tests of Mental

Table 1

Chi Squares for the P-F, Social, and Otis Tests

Variable	*df*	P-F Test Categories							Otis test
		Extrapunitive	Intropunitive	Impunitive	Obstacle-dominance	Ego-defense	Need-persistence	Denial	
Social intelligence									
Judgment in social situations	1	.47	−.08	2.00	.16	.16	−2.33	.04	.02
Recognition of mental state of speaker	2	−11.30***	4.85*	7.48**	2.02	−.94	.76	−.67	3.10
Memory for names, faces	2	−3.38	6.59**	−.20	−1.80	−.29	−1.23	20.80***	.01
Observation of human behavior	1	.63	−1.40	−.41	.00	−.29	1.11	−.41	.04
Sense of humor	2	.40	−1.54	−.17	.09	1.14	−1.65	.09	.90
Total	1	−.99	.23	.22	.08	−.15	.00	−.09	.08
Otis test	1	.36	.36	−.36	11.00***	−.36	22.60***	−.70	

Note—Negative sign indicates that the trend of the relationship is inverse.
* Significant between .05 and .10 level.
** Significant between .01 and .05 level.
*** Significant beyond .01 level.

Ability, Higher Examination, Form D [2]. This well-known test consists of 75 heterogeneous items relating to such aspects of general intelligence as measured by number series, analogies, reasoning, etc. The shorter 20-minute time limit was used with this test.

Results and Discussion

The major results are summarized in Table 1 which shows the interrelationships among the scoring categories of the Rosenzweig P-F test, the subtests of the Social Intelligence test, and the Otis Intelligence test. The entries on this table are chi squares determined from contingency tables in which the scoring categories of the P-F test were divided into two categories by median split while some of the subtest scores of the Social Intelligence test were divided by median split and others into high, middle, and low categories. The latter type of division was necessary because in some of these subtests too many scores were concentrated at the median. Thus some of the responses are based on 2 × 2 and some on 3 × 2 contingency tables.

Inspection of these tables discloses that most of the values are not significant. However, there are more highly significant chi squares than would be expected in a random sample of such chi squares.

The salient feature of this table lies in the fact that the major scoring categories of the P-F test (Extrapunitive, Intropunitive, Impunitive) yield two highly significant chi squares and one near-significant chi square with the subtest entitled Recognition of the Mental State of the Speaker (RMSS) on the Social Intelligence test. To evaluate these relationships let us first describe briefly the types of items in this subtest and in the P-F test. The subtest RMSS consists of eighteen 4-choice multiple-choice items in which the alternative that best describes the mood of a given quotation has to be selected. One item, Number 16, is as follows:

> "If only I had not let that opportunity slip I might now be enjoying the things I had longed for."
>
> (A) Despair (B) Disappointment (C) Loneliness (D) Regret

A comparison of this type of item with the items in the P-F test suggests a basic similarity in that the P-F test item presents a situation which provides information that may be used to characterize the mental state of one of the speakers. For example, in Item 2 there is shown a situation in which fragments of a vase lay scattered on the floor near a table and a woman is seen saying to another woman "How awful! That was my mother's favorite vase you broke." Obviously the testee provides the response for the other person. It is clear that the statement as made in the context depicted by the cartoon provides a basis for assessing the mental state of the speaker. It should be noted that this type of item provides the information about the mental state of the speaker on the basis of *both* verbal and pictorial material as against only the verbal material provided by the Social Intelligence test.

The question with which we are specifically concerned is whether the ability to recognize the mental state of the speaker as measured by the Social Intelligence test is importantly involved in the response made to the P-F test type of item. The aforementioned significant and near-significant chi-square values between the main scoring categories of the P-F test and this subtest suggest an affirmative answer to this question. This conclusion is further documented by the fact that a fairly clear and specific rationale can be found for these relationships.

Thus, the Extrapunitive (*E*) category yields a chi-square value significant at better than the .01 level. In addition, the relationship is inverse in the sense that a high *E* score tends to go with a low RMSS score. The meaning of this result seems to be fairly straightforward. The high-*E* individual is typically very outwardly aggressive and tends not to modify his responses in accordance with the needs of others. If he did, it would necessarily limit the outward expression of his aggression. It should be cautioned at this point that there are other dimensions (*N-P*, *O-D*, etc.) in the description of the *E*-type person. We have here emphasized the most important and pertinent dimension for the purpose of this comparison.

The second most significant chi square obtained is between the RMSS category of the Social Intelligence test and the P-F test Impunitive (*M*) scoring category. The high-*M* individual tends to be fairly objective in his reactions to situations in that he does not react in an aggressive manner either outwardly or inwardly; he does not, in general, place blame for the difficulty and tries to minimize the frustrating features of the situation. This type of behavior may be due to a lack of concern for the situation (in extreme cases, a type of schizoid reaction), or it may be due to a desire to protect the feelings of all involved. It is this last characteristic, we feel, that provides the basis for the positive relationship between the RMSS category and this one on the assumption that the desire to protect the feelings of others arises from an increased awareness of and sensitivity to their mental states.

Finally, the Intropunitive (*I*) category of the P-F test displays a tendency to vary positively with the RMSS score. Chi square is significant here between the .05 and .10 levels so that by conventional statistical standards the validity of this relationship is more speculative than were the other two. However, it is not difficult to find a rationale for this positive association. The high-*I* person is often one who tends first to blame himself when involved with frustrating problems. We may hypothesize that a basis for this self-blame lies partly in a developed appreciation and exaggerated sensitivity to other people's reactions.

The other Social Intelligence test scoring category to show significant relationship to the P-F test is Memory for Names and Faces (MNF). At the outset of the test the subjects are required to study for four minutes the photographs of the faces of twelve college-age males with a name below each face. Midway in the test the subjects are shown twenty-five faces among which are distributed the twelve originally examined. For each of the twelve names four of the twenty-five faces are suggested from which the subject must choose the correct face associated with the given name. MNF is significantly related to the Intropunitive (*I*) and Denial (*D*) scoring categories of the P-F test.

Examination of the two types of "items" subsumed under the categories of the respective tests reveals no palpable resemblances. In fact the facial characteristics of the cartoon figures in the P-F test are deliberately left unstructured to enhance its projective features (sex and age, however, are discernible).

Thus, unlike the comparisons discussed above, the form and content of these two types of items do not provide a ready rationale for explaining the high positive association indicated by the very significant chi squares obtained. On the basis of the conventional descriptions as well as our own experience, a rather speculative hypothesis to account for this association is offered. Specifically the Denial (*D*) category is symptomatic of generalized ego-insecurity. In fact, the high *D* is specifically suggestive of this characteristic. Our experience with people who are predominantly characterized by ego-insecurity suggests that these tend to be alerted to the facial expressions of others to a high degree. The reason for this is primarily due to their need for a continual assessment of their position in a social situation and this they must do by observing the behavior of others. The *D*-type person, being very anxious, tends to rehearse the events of previous situations including the observed behavior of people in these situations. This is a symptom which often goes with marked anxiety such as is created by insecurity. It is hypothesized, therefore, that the excessive attention to and rehearsal of others may have developed his ability to recognize faces. Thus, in the MNF situation at least a part of the task, that is, recognizing a face, is facilitated. Admittedly the empirical links between this hypothesis and performance on the two types of items are rather indirect. Still, it would not be impossible to test this hypothesis by correlating the MNF subtest with some other measure of feelings of inadequacy.

It may be noted also that the Intropunitive (*I*) category shows significant relationship with MNF ($<.05$). The writer can offer no basis for this other than to suggest the possibility that the *I* and *D* categories may be correlated.

Turning now to the interrelationships between the Otis Test of Mental Ability and the P-F test, we find two striking results, namely, highly significant chi squares for the Obstacle-Dominance (*O-D*) and Need-Persistence (*N-P*) categories. High score in the *O-D* category indicates, other things equal, an individual who emphasizes the frustrating character of the situation, that is, the particular circumstance which is at the root of the frustration. Thus, he is highly task-oriented. It may be that the significant chi square in this instance indicates that the more intelligent subject tends habitually to emphasize the problem aspect of the situation rather than the personalities involved in the situation. In short, it may be that the more intelligent individual, because of his greater success with problems, exhibits a tendency to be task-oriented rather than ego-oriented. A similar rationale may be employed in the case of the *N-P* category. In fact, to an even greater extent this type of individual is concerned with solving the problem involved in the P-F test.

Summary and Conclusion

The present study compared performance on three tests, the Rosenzweig P-F, George Washington University Social Intelligence, and the Otis Mental Ability tests. The purpose of this comparison was, in part, to determine the relationship between the influence of such abilities as are measured in the two types of intelligence tests on performance in the P-F test. The significant relationships which were found suggest the possibility that certain ability factors are related to the type of response made on the P-F test. This finding also suggests the general hypothesis that differences in the way the P-F test situations are perceived partly condition responses made to the test.

Received November 10, 1952.

References

1. Moss, F. A. The measurement of social intelligence. *Psychol. Bull.*, 1928, **25**, 168–169.
2. Otis, A. S. *Otis self-administering tests of mental ability* (manual). Yonkers, N. Y.: World Book Co., 1928.
3. Rosenzweig, S. The picture-association method and its application in a study of reactions to frustration. *J. Pers.*, 1945, **14**, 3–23.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

The Practical Utility of the Clinical Report

Charles A. Dailey
Veterans Administration[1]

This paper will present a technique of measuring the practical utility of a clinical report. It is an attempt to find an objective way of answering the question, "How should patients or clients be described in the psychological report?"

A recent empirical study by Davenport [3] dealt with this problem from the standpoint of the relationship of the clinical report to the data upon which it is based. In contrast, the present technique is concerned with the relationship of the clinical report to the treatment process itself.

For purposes of the present discussion, the *utility* of a report is defined as relevance to possible decisions about treatment of the patient described in the report. In measuring utility, one must compare the report with the treatment processes employed in the institution in which the report is to be used.

Since a report is always made *to* someone, semantic *clarity* is also involved in appraising its practical value. It is essential to study clarity for the reason that it is possible to say something relevant about a patient without saying it clearly.

The Measurement of Utility

The present method of measuring utility requires the following steps.

1. We first obtain a list of the decisions which are ordinarily made by the particular counseling, neuropsychiatric, or psychological service. In the exploratory study described here, we requested the psychologists, psychiatrists, and social workers of our service to recommend decisions about treatment which, in their opinion, are frequently made concerning the patients of their service. After refining these recommendations, we obtained the list shown in Table 1. It will be noted that this list contains a number of decisions which probably have no relevance to psychological reports. They are nevertheless included for the purpose of illustrating the technique. It is also obvious that the list is only a crude approximation to the decision-making process as it actually operates on such a service.

2. As a second step, the list was submitted to a group of ten clinical psychologists familiar with this service. Each was asked to "make the set of 32 decisions as if you were dealing with the *average* neuropsychiatric patient on this service. Indicate your opinion as to the appropriate treatment for this hypothetical average patient by stating 'yes,' 'no,' or 'don't know' in response to each question." The responses of these clinicians were compiled and are reported in Table 2. This consensus of clinical opinion provides a base line from which the new information value of a report may be calculated. Such a measure is important to provide a control upon stereotypy in decision making; we must make certain that, in making decisions about a patient after reading a report about him, the clinician is not merely repeating a priori decisions.[2]

3. The third step requires the appraisal of the reports which are to be measured for utility. A clinician who is otherwise unfamiliar with the specific patient reads each report and then fills out, in checklist form, his recommendations for each of the decisions about treatment. His recommendations are then scored as to their deviation from the a priori recommendations (Table 2). A weighting system is employed for this purpose which attributes to each decision the inverse of the proportion of clini-

---

[1] From the Veterans Administration Hospital, Jefferson Barracks, Missouri.

[2] The author is indebted to the work of Cronbach [1] and of Kelly and Fiske [5] for emphasis upon the concept of relative information value in psychological appraisal.

Table 1

List of Frequently Made Clinical Decisions

1. Should the patient have been admitted to the psychiatric ward?
2. Should he be put on closed ward (for reasons other than electroconvulsive therapy)?
3. Should he receive electroconvulsive therapy?
4. Should he receive subshock insulin?
5. Should he receive intensive, interpretive individual psychotherapy?
6. Should he receive supportive, noninterpretive psychotherapy?
7. Should he receive group psychotherapy?
8. Should he be restricted for the first three days?
9. Should he receive repressive individual psychotherapy (i.e., therapy directed toward inhibiting completely some impulses and ideas)?
10. Should he be allowed visitors now from immediate family?
11. Should he receive sedative drugs?
12. Should he receive vitamins?
13. Should he be placed on a special diet?
14. Should he receive placebo?
15. Should he have amytal interviews for cathartic purposes?
16. Should he have hypnotic treatment for cathartic purposes?
17. Should he have a male therapist?
18. Should the therapist play an authoritarian role?
19. Should the therapist play a nurturant role (i.e., kindly and especially tolerant of dependency needs)?
20. Should the patient have occupational therapy?
21. Should this patient discuss his vocational future with the vocational counselor?
22. Should the patient work in Manual Arts Therapy?
23. Should the patient be transferred now to a mental hygiene clinic?
24. Should his relatives be given therapeutic interviews?
25. Should the patient receive especially solicitous attention from ward personnel?
26. Should the patient be particularly encouraged about his resources for "adequate living," "success," etc.?
27. Should the patient be watched for assaultive tendencies?
28. Should the patient be watched for suicidal tendencies?
29. Should the patient be encouraged to make his own decisions about length of stay?
30. Should the patient be encouraged to improve his educational status?
31. Should the patient be encouraged to change his job?
32. Should the patient be encouraged to seek the aid of Alcoholics Anonymous?

Table 2

Responses of Ten Clinicians to the Question of What Treatment Should be Given to the Average Neuropsychiatric Patient

Treatment decision number	Number of clinicians responding Yes	?	No
1	9	1	0
2	0	4	6
3	0	5	5
4	0	6	4
5	2	3	5
6	8	2	0
7	9	0	1
8	9	1	0
9	1	4	5
10	7	2	1
11	2	6	2
12	6	3	1
13	1	5	4
14	2	5	3
15	0	3	7
16	0	2	8
17	6	4	0
18	0	5	5
19	4	5	1
20	10	0	0
21	6	2	2
22	8	2	0
23	1	6	3
24	7	2	1
25	6	1	3
26	6	1	3
27	4	2	4
28	3	2	5
29	7	2	1
30	0	4	6
31	0	3	7
32	0	5	5

cians recommending that treatment for the average patient. To derive this system, each number in Table 2 is subtracted from ten; the remainders are a set of weights. The scores are summed for the 32 decisions; the resulting total score represents the tendency of the report to suggest to the clinician decisions which would not have been made on the basis of "zero" information about the patient.

4. If the total score is divided by ten (no transformation is necessary if proportions are used in Step 3, instead of raw number of clinicians, to derive the weights), the resulting number represents the number of the 32 decisions which were "completely novel." This is, of course, a hypothetical number which assumes

that two "half-novel" decisions should be equated to one "completely novel" decision (i. e., one which is made exactly opposite to the a priori expectation). This variable allows us to compare various weighting systems and to make full use of our experimental data.

We shall hereinafter refer to this measure as *number of new, useful decisions.* Table 3 presents a distribution of such measures, in which the decisions of two judges for each patient are used. Since two judges are used, this modification of the measure is called "number of *clear,* new, useful decisions," for reasons which will be elaborated in the following sections.

The Measurement of Clarity

The clarity of a report is defined as the amount of agreement of two independent readers. As we are concerned with utility, it is understood that the agreement is to be about *treatment* (as distinguished from, for example, the more esoteric categories used in studying reliability of personality judgments). Clarity is appraised in the following steps.

1. Two clinicians read a given report independently. The report constitutes their sole information about the patient.

2. Following the reading, they are requested to fill out the decision check list (Table 1) independently, responding "Yes," "No," or "Don't know" with regard to each decision.

3. Their decisions are compared and a score, *number of clear, useful decisions* (the number upon which they agree), is calculated. A distribution of such scores is presented in Table 3.

The measurement of clear, new, useful decisions. It is desirable to compute an over-all value which combines the opinions of separate judges, and which also describes the departure of their opinions from a priori convictions about treatment.

For this purpose, the weighting system described above (under *The Measurement of Utility*) is applied to those decisions upon which two clinicians separately concur. The resulting measure is termed *number of clear, new, useful decisions.*

The Clarity and Utility of Nine Psychological Reports

Nine clinical reports were selected at random from our files. The reports were distributed to seven clinicians familiar with our treatment procedures. They were distributed in such a way that each report was read by two clinicians and so that no two clinicians read more than one report in common.

Table 3

Clinical Utility and Clarity in Nine Reports

Patient	Number of clear, useful decisions	Number of new, clear, useful decisions	Proportion of clear decisions that are new
A	19	9.4	.50
B	18	10.8	.60
C	15	7.8	.52
D	22	12.0	.54
E	14	8.3	.60
F	11	5.2	.47
G	11	5.1	.46
H	22	9.5	.43
I	17	7.2	.42
Median	17	8.3	.50
Per Cent of Total*	.53	.26	

* The total is 32 decisions.

On concluding his study of the report, each clinician was requested to make the 32 decisions (Table 1) for the patient described in the report. The responses were analyzed according to the operations described above. Results are summarized in Table 3. In interpreting Table 3, we can state that slightly over half of the 32 decisions can be clearly agreed on after reading the report. Of those which can be clearly agreed on, only half represent decisions which would not have been made before reading the report (according to Table 2).

Thus, about one fourth (26%) of the list of therapeutic decisions are clearly indicated contrary to a priori expectations in the average psychological report, within this sample of reports. It is, of course, not possible to say whether this is a "satisfactory" or an "unsatisfactory" proportion, with our present level of knowledge about the proper use of these treatments.

We can, however, go on to consider the

problem of which features of the psychological report may influence its clarity and utility.

Linguistic Correlates of Clarity and Utility[3]

The clinical reports which were measured above as to clarity and utility were also measured as to several factors which describe style and properties of the writing in each report:

1. length of report (in words)
2. average sentence length (in words)
3. average number of technical terms (per word)
4. average number of personalizing terms (per word)

The variables "technical terms" and "personalizing terms" require definition. The author selected from a large number of psychological reports those words which appeared to be frequently found in the vocabulary of psychologists but not in the conventional speech of lay persons (e. g., "dysfunction"), or which, while appearing in the speech of lay persons, appear to receive meanings different from the meanings assigned them by psychologists (e. g., "deficit"). This list contains 140 words and its importance does not justify its inclusion in the present paper.

Frequency of technical words is defined as the total number of words from such a list divided by the total number of words in the report. Such a frequency count was made in each of the nine psychological reports; the distribution of values is summarized in Table 4.

"Personalizing terms" include all personal pronouns (excluding "it"), include all nouns that have a masculine or feminine natural gender (e.g., "nurse") but not common gender nouns (e.g., not "authority"), and include plural nouns with a group referent (e.g., "folks"). This definition of personalizing terms parallels the definition offered by Flesch [4] in his readability studies.

Frequency of personalizing terms is therefore defined as the total number of personalizing terms divided by the total number of words in the report. Distribution of these values for the nine reports studied is presented in Table 4.

Each of these four variables: length of report, length of sentence, frequency of technical terms, and frequency of personalizing terms, was correlated by rank-order method with the variables pertaining to utility and clarity of the report. Results of these correlations are presented in Table 5. This table also presents reliability figures for the utility and clarity variables.

Table 4

Linguistic Properties of Nine Psychological Reports

Patient	Report length (words)	Sentence length (words)	Technical terms (per word)	Personal terms (per word)
A	730	19.7	.20	.13
B	359	15.6	.22	.08
C	467	20.3	.22	.05
D	595	20.0	.22	.07
E	254	16.9	.17	.04
F	362	17.2	.18	.05
G	430	17.9	.17	.08
H	604	18.3	.22	.06
I	382	22.5	.23	.05
Median	430	18.3	.22	.06

Table 5

Rank-Order Correlations between Clarity, Utility, and Linguistic Variables

($N=9$ reports, each judged by two clinicians)

Linguistic variable	Number of clear, useful decisions	Number of new, clear, useful decisions
Report length	.66	.27
Sentence length	−.10	−.15
Technical terms	.64	.43
Personal terms	.36	.32
Number of clear, useful decisions	.59*	
Number of new, clear, useful decisions		.58*

* These correlations are reliability coefficients calculated by the split-half method.

A Priori Likelihood of a Decision

Because of the uniformly low rank-order correlations involving the variable, "number of new, clear, useful decisions" (Table 5), it is appropriate to question the assumption that the weighting system used in measuring this variable has any actual relationship to the process of clinical decision. Are the recommendations for therapy made by clinicians when they are given *no* information about a patient (Table 2) similar to the recommendations which they make after obtaining information about a specific patient?

---

[3] The author is indebted to M. Buttiglieri and R. Phillips for their analysis of these linguistic variables.

The author demonstrated in previous experiments [2] that the stating of an opinion about a person led a judge to "fixate" his opinion. As a further question about the nature of clinical "rigidity," the similarity between clinical judgments before and after reading a report should be measured.

Both these questions (the reliability of the weighting system and the rigidity of clinical recommendations) converge in the data presented in Table 6. This table compares the amount of agreement, among ten clinicians, in recommending each of the 32 treatment procedures for the "hypothetical average N-P patient," with the amount of agreement among those same clinicians in recommending the 32 treatment decisions for the specific patients whose reports they read. Thus, each treatment decision receives two agreement scores. Table 6 compares the treatment recommendations "before" and "after" reading reports.

Table 6

Correlation between Decisions Made about the "Average" Patient with Decisions Made about 9 Specific Patients

Number of clinicians making a given decision for the hypothetical average N-P patient *	Number of times same decision made regarding any specific patient after his psychological report read by a clinician †		
	0–3	4–9	10–18
0	7	2	0
1–6	3	8	3
7–10	1	2	6

$N = 32$ decisions
Chi square $= 17.4$‡
$df = 4$
Significance less than .01

* Ten clinicians made these decisions for the "hypothetical, average N-P patient" (Table 2), making a total of ten sets of decisions.

† Nine reports were each read by two clinicians; each clinician made a set of decisions for each report that he read, making a total of eighteen sets of decisions.

‡ The Pearsonian r for these data, when they are ungrouped, is .76. This is a highly significant value.

Table 6 demonstrates that the weighting system used in measuring the variable, *number of new, clear, useful decisions*, does seem to have a relationship to the process of clinical decision about what treatment should be given to *specific* patients. Whether this also demonstrates the strong influence of clinical bias upon treatment planning remains to be determined by further experimentation. The finding could also be interpreted as indicating that the information on which reports are based (i.e., psychological testing) is poorly developed in the direction of treatment planning.

It seems safe to conclude (*a*) that the weighting system has some reliability and (*b*) that the factor of clinical bias (among psychologists, at least), or stereotypy of opinion, helps determine the role of the report in the treatment-planning process. Such conclusions have interest in relation to the method of measurement described in this paper.

Conclusions Regarding the Measurement of Utility and Clarity

1. The number of clear, useful decisions implied by a psychological report can be measured by counting the number of decisions agreed upon by two clinicians who have read the report independently. This measure has a split-half reliability of .59 in the small sample of nine reports studied.

2. The number of *new*, clear, useful decisions implied by a psychological report can be measured by weighting each decision in (1) above, according to the inverse of the likelihood that such a decision could be made with reference to the average patient. This measure has a split-half reliability of .58 in the small sample studied.

3. The weighting procedure employed in (2) appears to be quite reliable in that clinicians tended to make the same therapeutic decisions after reading psychological reports that they had previously made for the "average N-P patient." This similarity between a priori and a posteriori recommendations may explain why greater relationships between the "readability" and other linguistic variables on the one hand, and, on the other hand, the variable *new, clear, useful information*, were not obtained. The operation of clinical bias or rigidity may be so great as to mask many of the effects of the psychological report.

4. Suggestive but not convincing relationships between the number of clear, useful decisions and the length and technical content of the report were found. The longer a report

(between the values of 254 and 730 words), the greater the clarity in terms of therapeutic implications; and, the more technical the terminology (independent of length), the greater the clarity.

Implications for Further Research

The above data and findings are presented only for illustrative purposes. The intent of this study is primarily methodological.

The reliabilities of the measures of utility and clarity are high enough to be promising but low enough to require further refinement of the measures before large-scale inquiry is justified.

Limitations on the general use of such measures are imposed by the fact that, logically, such measures have to be devised with reference to the institutions within which they are to be used. Variation in institutional lists of decisions and particularly in a priori weights (i.e., clinical premises) would reduce the general applicability of any one scale.

It should also be emphasized that the present method defines only one dimension among the many along which the value of a psychological report should be calculated. A report has relevance not only to gross decisions such as those listed here, but also to subtler problems which arise during protracted psychotherapy. Other purposes of the report—such as establishing in the therapist or decision maker a clear and memorable image of the patient—are not met by the present method.

Nor does the method help solve the problem of whether *valid* decisions have been suggested by the report. The broader problem of whether utility and clarity have any relation to validity remains for future research.

The most important suggestion of the present discussion is that the "action-potential" of a report should be a criterion of its usefulness; the most important implication of our data is that it is possible to measure this property. It is possible that we may uncover empirical criteria which will ultimately define the language which should be spoken by the clinician.

Received November 6, 1952.

References

1. Cronbach, L. J. A generalized psychometric theory based on information measure. (Mimeographed paper.) Urbana, Ill.: Bureau of Research and Service, Univer. of Illinois, 1952.
2. Dailey, C. A. The effects of premature conclusion upon the acquisition of understanding of a person. *J. Psychol.*, 1952, **33**, 133–152.
3. Davenport, Beverly F. The semantic validity of TAT interpretations. *J. consult. Psychol.*, 1952, **16**, 171–175.
4. Flesch, R. *The art of readable writing.* New York: Harper, 1949.
5. Kelly, E. L., & Fiske, D. W. *The prediction of performance in clinical psychology.* Ann Arbor, Mich.: Univer. of Michigan Press, 1951.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Clinical Judgment and the Cattell Infant Intelligence Scale

James J. Gallagher
Michigan State College

It is a maxim in psychology that the efficiency of clinical practice is determined by the validity of the clinical judgment of the individual psychologist. Although called upon to use this judgment every day in practice, the average psychologist has had little opportunity to check systematically the validity of his own judgment because of pressures of business and some difficult methodological problems. One of the many situations in which the clinician must use his clinical judgment is when he is asked to test infants and give an opinion as to their intelligence, oftentimes at the request of an adoption agency. Good judgment is especially needed in this situation since, for obvious reasons, test conditions are often far from ideal when infants are being evaluated. Every clinician has in his or her reports written some such comment as:

> "This child did not appear to be doing his best."
> "The child was so irritable that the examiner feels her performance was lower than it otherwise would have been."
> "The infant appeared sluggish and sleepy and did not respond to many of the items."

Yet the opportunity of checking these statements comes only too seldom to the busy clinician. The purpose of the present study was to determine if there were any significant changes upon retest in a group of children who were judged as not doing their best upon the original infant test.

Previous Studies

The entire area of infant testing has been a neglected one in psychology. Although both the Gesell Developmental Scales [3] and the Cattell Infant Intelligence Scale [1] have been in common use for some time, little has been reported concerning their validity and reliability beyond the original presentation of the authors. A new scale designed for infants from 4-12 weeks has recently been developed at Northwestern University [4] but is not yet in general clinical use. Watson [6] has recently stated that no published study, other than that of Cattell herself, concerning the scale was known, although it was then ten years old. Since then Escalona [2] has done a study on the use of infant tests for predictive purposes. Using both the Cattell and Gesell scales she found that a group of children tested under nonoptimal conditions tended to vary more on retests than did a group of infants who were tested under what was judged to be optimal testing conditions. The measure of change was, in this case, movement in broad intellectual classifications, i.e., such as from Dull Normal to Average. Even the group tested under optimal conditions showed a marked tendency to vary one intellectual degree away from the original test in 31% of the cases. Her conclusions were that for the best predictive results the gestalt of the testing situation had to be taken into account rather than the mere test finding itself.

Method

The subjects in the present study were forty-three infants, ranging in age from 4.1 months to 24.1 months, who were seen at the Dayton Guidance Center over a period of three years. All these infants had been referred by community agencies supervising adoption cases, and all had taken the Cattell Scale at least twice. Since the general policy of the psychologists was that retests of all infants were desirable, it was possible for the writer to divide the children into two groups: those having routine retest and those who were suspected of not doing their best on the original tests. The two psychologists who did the tests had M.A.'s in psychology plus several years of experience.

The writer, who did none of the actual testing, chose the children for the Mandatory Retest group upon the following criteria:

a. Evidence in the case history of extremely deprived environmental background up until the present examination (such as malnutrition, desertion by parents, etc.).

b. Definite statements in the psychological report that owing to poor rapport conditions (child crying — restless — not responding) the psychologist believed retesting necessary.

The children placed in the Routine Retest group were selected on the basis of the absence in the psychological reports of any comments indicating special problems existing in the test situation or in their previous environment. The placement of each case in its proper group was quite clear owing to the precise wording of the reports. A recheck of the placement of twenty-five of the cases, chosen at random and read two months after the original selection by the writer, found twenty-three of the twenty-five cases being placed in the same category as before. The mean age of Routine Retest group at the time of first testing was 8.29 months, while the mean age for the Mandatory Retest group was 10.53. This was not a statistically significant difference. The mean time interval between the first and second test for the Routine group was 7.90 months and for the Mandatory group was 7.18 months.

Table 1

Test-Retest IQ Scores on Mandatory and Routine Retest Groups

Measure	Mandatory Retest	Routine Retest
First test		
N	19	24
Mean	88.05	100.62
SD	17.80	9.18
Retest		
N	19	24
Mean	96.58*	101.25
SD	14.79	8.99
Test-Retest correlation	.77	.83†

* Difference between means significant at the 1% level of confidence.

† Corrected for restriction of range [5].

Table 2

Changes in Terman-Merrill Classifications As Result of Retest

Test	Two below	One below	Same category	One above	Two above	Three above
Routine Retest		4	16	4		
Mandatory Retest			11	7		1

The standard *t* test was used to determine the significance of difference between the pre- and posttest mean IQ's of each group. The changes in the standard Terman-Merrill [5] classifications as a result of the retest were also calculated.

Results

Table 1 shows that the Mandatory Retest group made a mean gain of 8.53 IQ points at the time of retest, and the difference between first and second mean IQ scores was significant at the 1% level of confidence. The Routine Retest group showed a gain of .53 IQ points on the retest, a difference that was not statistically significant. The test-retest correlation for the Mandatory group shows a respectable but not outstanding *r* of .77. The Routine Retest group showed a test-retest *r* of .52 but when this is corrected for restriction of range, this *r* becomes .83.

Table 2 reveals the change in Terman-Merrill classification as a result of the retest of both Routine and Mandatory Retest groups. In the Routine group the differences were evenly split, four infants dropping one classification while four moved up one classification. These results closely agreed with Escalona's findings of her optimal test group, with 31% of her cases falling in an adjacent range, as compared with 33% of the present study. In the Mandatory group it may be seen that the changes are all in the direction of a higher retest classification and one infant even changed to three classifications higher.

Conclusions

1. In this particular situation it appears that the clinical judgment of the psychologists was sound in stating as necessary further testing for those infants who did not appear to be doing their best. While these results are limited to the situation involved in the study, it is

encouraging to note the similarity of the results in this study and in Escalona's study.

2. It can be noted that although the results of the Routine Retest group show them to have a much more stable retest pattern than the Mandatory Retest group, their own reliability cannot be considered too good for practical circumstances. All doubtful cases should be seen again rather than risk damage to a prospective adoption case by too much reliance on a useful but not overly reliable instrument.

3. Evidence seems to support the conclusion that all possible information concerning not only test results but the conditions surrounding the test and full case histories are necessary for adequate competence in the prediction of infant intelligence.

Received October 30, 1952.

References

1. Cattell, Psyche. *The measurement of intelligence of infants and young children.* New York: Psychological Corporation, 1940.
2. Escalona, Sibylle. The use of infant tests for predictive purposes. *Bull. Menninger Clin.*, 1950, **14**, 117–128.
3. Gesell, A., & Amatruda, Catherine. *Developmental diagnosis.* New York: Hoeber, 1941.
4. Gilliland, A. R. *Northwestern intelligence tests. Test A. Test for infants 4–12 weeks old.* Boston: Houghton Mifflin, 1949.
5. Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1950.
6. Terman, L. M., & Merrill, Maude A. *The measurement of intelligence.* Boston: Houghton Mifflin, 1937.
7. Watson, R. I. *The clinical method in psychology.* New York: Harper, 1951.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Three Factors in Parent Behavior

Maurice Lorr and Richard L. Jenkins
Veterans Administration[1]

The importance of the family and the role of the parents in the personality development of the child have been widely recognized. Much clinical evidence has accumulated with regard to the influence of parent attitudes, and to the effect of various methods of handling, on the child's emotional development. However, until recently, comparatively little systematic effort has been directed toward the identification of basic variables in parent behavior.

The Fels Parent Behavior scales developed by Champney [2] represent an effort to categorize and measure a wide range of family variables. The variables covered in the 30 rating scales were intended to measure eight basic "armchair" factors of parent behavior. The eight factors or areas describe the following: home atmosphere, contact of child and mother, babying and protectiveness, criticism and evaluation of the child, readiness of explanation, emotional relationships between parent and child, and a miscellaneous group.

Baldwin, *et al.* [1], using the rating schedule developed by Champney, secured ratings by home visitors on factors in each home in a population whose children were also being studied in other ways. Subsequently, in an effort to secure simplification of the 30 variables, Roff [3] isolated seven common factors by means of a multiple factor analysis of the intercorrelations between the Fels Parent Behavior scales obtained by Baldwin.

Table 1

Matrix of Correlations of Primary Factors

		I	II	III	IV	V	VI	VII
Concern for child	I	1.00						
Nonreadiness of suggestion	II	–.36	1.00					
Democratic guidance	III	.37	.06	1.00				
Parent-Child harmony	IV	.38	.11	.63	1.00			
Nonpermissiveness	V	–.36	.14	–.34	–.36	1.00		
Activeness of home	VI	.03	–.11	.27	.04	.36	1.00	
Sociability of parents	VII	.23	.26	.32	.06	.22	.26	1.00

Table 2

Rotated Oblique Second-Order Factor Matrix r_{tg}
and Transformation Matrix ψ_{jg}

			r_{tg}			ψ_{jg}	
Rating Scales		X	Y	Z	X	Y	Z
Concern for child	I	.781	.012	.092	1.071	.016	.126
Nonreadiness of suggestion	II	.578	–.338	–.012	–.732	.428	.015
Democratic guidance	III	.022	.770	–.001	.034	1.197	–.002
Parent-Child harmony	IV	.011	.592	–.236	.014	.735	–.293
Nonpermissiveness	V	–.068	–.448	.601	.096	.633	–.849
Activeness of home	VI	.134	–.153	.569	.174	–.198	.734
Sociability of parents	VII	.001	–.005	.440	.001	–.007	.602

---

[1] From Veterans Administration Central Office, Washington, D.C.

Table 3

Factor Loadings of Rating Scales on Second-Order Factors

Rating Scales	*X*	*Y*	*Z*
Babying: overhelps — withholds help	.651	–.006	.054
Child-centeredness of home: child-centered — child-subordinated	.640	–.025	.078
Solicitousness for child's welfare: anxious — nonchalant	.622	.133	.035
Protectiveness: sheltering — exposing	.600	.023	.245
Intensity of contact: vigorous — inert	.576	.035	.060
Duration of contact with parent: extensive contact — brief contact	.556	.054	.019
Acceptance of child: devotion — rejection	.438	.280	.015
Readiness of suggestion: suggesting — nonsuggesting	.385	–.294	.061
Democracy of policy: democratic — dictatorial	–.039	.764	–.087
Noncoerciveness of suggestion: optional — mandatory	–.029	.735	–.033
Justification of policy: rational — arbitrary	.019	.696	.030
Nonrestrictiveness of regulations: freedom — restriction	–.294	.661	–.262
Nonemotionality toward child: objective — emotional	–.204	.625	–.211
Nonreadiness of criticism: uncritical — critical	–.263	.617	–.327
Readiness of explanation: satisfies curiosity — thwarts curiosity	.060	.586	.198
Favorableness of criticism: approval — disapproval	.199	.572	.004
Understanding of child's problems: keen — obtuse	–.027	.490	.256
Clarity of policy: clear — vague	.001	.461	.262
Nondisciplinary friction: concordant — contentious	.080	.441	–.145
Affectionate toward child: affectionate — hostile	.321	.387	–.130
Rapport with child: close rapport — isolation	.312	.347	–.128
Nondiscord in home: harmony — conflict	–.019	.328	.124
Effectiveness of policy: successful — unsuccessful	.108	.299	–.063
Readiness of enforcement: vigilant — lax	.016	–.318	.592
Severity of actual penalties: severe — mild	–.079	–.425	.473
Accelerational attempt (pushing): acceleratory — retardatory	.123	.229	.427
Coordination of household: coordinated — chaotic	.079	.074	.407
Activeness of home: active — inactive	.153	–.138	.394
Adjustment of home: well-adjusted — maladjusted	.030	.098	.369
Sociability of family: expansive — reclusive	–.056	–.044	.288

The factors isolated were interpreted and described as concern for the child, democratic guidance, permissiveness, parent-child harmony, sociability adjustment of parents, activeness of home, and nonreadiness of suggestion. The correlations between these primary factors as determined by Roff [3] are shown in Table 1. Examination of the table reveals substantial correlations between these factors. It appeared that an analysis of these correlations should reveal one or more underlying second-order factors of a broad character which should be interesting in themselves and useful for the conceptualization of parent-child relations. Accordingly, the intercorrelations in Table 1 were subjected to a centroid analysis and three orthogonal factors were extracted and rotated to oblique simple structure.

Table 2 presents the loadings of the first-order factors on the second-order factors. It also gives the transformation matrix for computing the factor loadings of the original 30 rating scales on the second-order factors [4].

It will be noted that the significant loadings on Factor *X* are from concern for the child and nonreadiness of suggestion.

The variables with the largest loadings on factor *Y* are democratic guidance, parent-child harmony, and permissiveness. Nonreadiness of suggestion also has a sizable negative loading.

The heaviest loading on factor *Z* is a loading from nonpermissiveness. Activeness of the home and sociability of the home also show large loadings.

The character of these second-order factors becomes much clearer when we refer to Table 3, which presents the correlations of the original 30 family characteristics with the second-order factors. The largest correlations with factor *X* are babying, child-centeredness of the home, solicitousness for the child's welfare, protectiveness, intensity of contact, duration of contact, and acceptance of the child; devotion rather than rejection. It would appear that this factor might be characterized as the satisfaction and encouragement of the child's emotional dependence on the parent. We will call it dependence-encouraging.

The second factor, *Y*, is best described by democracy of policy, noncoerciveness of suggestion, nonrestrictiveness, nonemotionality toward the child, nonreadiness of criticism, readiness of explanation, favorableness of criticism, understanding of the child's problems, clarity of policy, and the absence of disciplinary friction. This is clearly a factor of democracy and democratic values in child training in contrast with undemocratic methods. This factor will be called democracy of child training.

The third factor, *Z*, is characterized by vigilance of enforcement, severity of actual penalties, accelerational attempt (pushing), and coordination of household. Activeness of the home, the adjustment of the home, and the sociability of the family show smaller correlations. This is a factor of organization and effectiveness of control in the home. It ranges from a strict orderliness at one extreme to a chaotic disorder at the other.

It would appear from this analysis that, if we are to seek in clinical work to conceptualize parent-child relations in shorthand fashion, the three most economical questions we may ask ourselves are:

1. How far does this home sustain and encourage dependence and how far does it deny satisfaction to dependence?
2. How far do its methods of child training reflect democratic practices and values, or to what extent are they authoritarian and undemocratic?
3. To what extent is there a strict orderliness in the home or to what extent is the home lax and unorganized?

We must recognize, of course, that these three questions, based on the second-order factors, would not yield us all the information we would obtain from ratings on the seven first-order factors, any more than the latter would give us all the information we would obtain from ratings on the 30 schedules from which they were derived. Busy clinicians, however, can hardly afford routinely to make ratings on the 30 schedules, unless they are pursuing this as a research project. We submit that in terms of our present knowledge the three foregoing questions embody more of the significant elements of the larger study than any other equally economical condensation.

The intercorrelations between the three second-order factors were also computed. The correlation of −.43 between factors *X* and *Y* imply that the home with democracy of policy tends not to be dependence-encouraging. The correlation of −.47 between factors *Y* and *Z* suggests that a democratic policy is related to a home that is less organized and less effective in the enforcement of its policies than an authoritarian home may be. Factors *X* and *Z* are independent, indicating that there is no significant relation between the encouragement of dependence on the one hand and effectiveness of control in the home on the other.

Summary

A second-order factor analysis of the interrelations of seven first-order factors derived by Roff from 30 Fels Parent Behavior Scales indicates that the most economical and significant condensation of the significant relations tapped by these scales can be represented in the following three questions:

1. How far does this home sustain and encourage dependence and how far does it deny satisfaction to dependence?
2. How far do its methods of child training reflect democratic practices and values, or to what extent are they authoritarian and undemocratic?
3. To what extent is there a strict orderliness in the home or to what extent is the home lax and unorganized?

Received November 6, 1952.

References

1. Baldwin, A. L., Kalhorn, Joan., & Breese, Fay H. Patterns of parent behavior. *Psychol. Monogr.*, 1945, 58, No. 3 (Whole No. 268).
2. Champney, H. The variables of parent behavior. *J. abnorm. soc. Psychol.*, 1941, 36, 525-542.
3. Roff, M. A factorial study of the Fels parent behavior scales. *Child Develpm.*, 1949, 20, 29-44.
4. Thurstone, L. L. *Multiple-factor analysis.* Chicago: Univer. of Chicago Press, 1947.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Follow-up Study of Ten-Year-Old Boys with Police Records

William W. Wattenberg and Frank Quiroz
Wayne University

The present study is an attempt to check more directly some inferences drawn from the findings of an investigation reported by Wattenberg [3]. Data relating to eleven-year-old boys who had charges filed against them with the Detroit police were collected. These indicated that many of the boys were victims of neither severe personality disorganization nor bad environmental pressures. Rather, their misconduct presumably could be accounted for as resulting from poor judgment, weak internal controls, or other immaturities which they could be expected to outgrow.

There are sound theoretical grounds for believing this could be the case. Several recent writers [1, 2] on the psychology of preadolescence maintain that this period of development is accompanied in American culture by personality attributes, temporary in nature, which might lead a normally developing youngster to engage in adult-prohibited actions. These would not necessarily be symptoms of serious personality maladjustments or character defect. Although such actions might bring a youngster to the attention of the police, they would not be predictive of delinquency in subsequent years.

Two specific trends attributed to preadolescence as a growth phase are: (*a*) a disorganization of the childish personality structure and (*b*) a strong attraction to peer groups which engage in a species of guerrilla warfare against adults. Either trend, if it exists, could easily bring a youngster into collision with the forces of law and order. Both, presumably being temporary, eventually wane. Delinquency resulting from either or both would therefore be expected to vanish.

If there is any truth in these speculations, it has very practical implications. At the present time, clinics serving youth are badly overloaded in most areas. A good portion of their load consists of children referred by courts or brought by parents who have been disturbed by juvenile misconduct. Now, if in this load there is a large proportion of boys and girls who would have recovered even if nothing had been done for them, the overworked clinics are wasting precious time and even more precious skill. If such should be the case, then an obvious next step would be to search intensively for some reasonably effective means of discovering which preadolescents are genuinely in need of therapy. If this could be done, then clinics could concentrate on those youngsters who would become delinquent or mentally ill if they received no treatment.

From the viewpoint of scientific research designed to evaluate methods of psychotherapy the possibility we have been discussing has grave implications. If among the populations receiving treatment there is a large proportion who would have improved in the course of time without therapy, statistics based on percentages of "success" become meaningless. As a minimum precaution research workers would have to make allowances for the existence of this factor.

The study [3] mentioned contained indirect evidence that preadolescent boys in trouble with the police included a large proportion who were relatively normal youngsters. However, the data were neither gathered nor ordered in such a way as to give an estimate of how large the "normal" group might be nor to verify the hunch that they would recover "spontaneously" from their current misbehavior. Accordingly, it was decided to take a sample of ten-year-olds who had police records and to make a follow-up study to discover how many of them did continue to be delinquent. In addition, an effort was made to discover what available facts gathered at the time of their first police contact distinguished between those who continued to get in trouble and those who later had no police records for a twelve-month period.

For the purposes of this study, we accepted the following hypotheses as representing the points to be tested:

Hypothesis No. 1: In a group of ten-year-olds with police records, only a small fraction will be found two years later to be continuing to act as delinquents.

Hypothesis No. 2: The group which continues to act as delinquents will be distinguished by having greater pressures in the form of bad family situations, poor school relations, and bad neighborhood influences.

Procedure

Use was made of the very comprehensive records kept by the Youth Bureau of the Detroit Police Department.[1] The records for 1948 were searched, and all those pertaining to boys ten years old at the time of their first contact were segregated. These records were then placed in order by the serial number assigned to the boys. Next, the records of all boys contacted by the police during 1950 were placed in serial number order. The two lists were checked and the original records of all boys among our original ten-year-olds who had police contacts in 1950 were set aside. We termed these boys "repeaters." The nonrepeaters, then, were boys who after a period of at least one year (but averaging eighteen months) had gone for a full twelve months without any complaint against them having been made to the police.

The year 1948 was chosen because at the time the routine handling of ten-year-olds was either to release them to their parents with a warning or to have them come in to visit police officers weekly for a series of check-up conferences. Only where the offense was very serious did the police officers file a petition with the Juvenile Court. At the time, the clinic attached to the court had a heavy backlog of cases. With rare exceptions, notably in the instance of sex offenses, the first offenders were put on probation without being seen by the clinic. Therefore, it is safe to say that only a tiny fraction of the 1948 ten-year-olds had received systematic psychotherapy.

To test the second hypothesis, the repeaters and nonrepeaters were compared on 50 items of information reported in the Youth Bureau "history sheet," filled out by selected police officers after a visit to the home, an interview with the boy, and conferences with school personnel. For each item a tabulation was prepared and tested for statistical significance by the chi-square technique. In these comparisons only the records made at the time of the first police contact were used.

Findings

In all, the 1948 files contained records of first contact with 207 ten-year-olds. In 1950, only 43 of these had any contact with the police. Thus, close to 80 per cent of the group had gone a full year without any reported relapse into actionable misconduct. These figures furnish strong support to the first hypothesis.

In the comparisons between repeaters and nonrepeaters, the null hypothesis could be rejected at the 1 per cent level of confidence for three tables; at the 2 per cent level, for an additional tabulation; and at the 5 per cent level, for four more. The chance expectation would call for 2.5 tabulations to meet this criterion. Thus, the number of statistically significant differences was 200 per cent greater than should have occurred by chance alone. However, these results are clouded by the fact that although classifications were combined, in most instances of apparent statistical significance, the chance expectation in the critical cells fell below five. For what they may be worth, however, the results will be reported.

Of the apparently significant tables, two concerned the parents. As shown in Table 1, among the few boys for whom parental alcoholism was reported, almost half were listed among the boys who were still running up police records two years later. Also, as shown in Table 2, the majority of boys whose parents were regarded as uncooperative by the police, were also repeaters. As indicated by the column, "Contribution to chi square," in both these tables, the cells yielding the highest contribution to the chi-square table had chance expectations below five. Thus, the results can be considered suggestive, but not conclusive.

---

[1] The authors wish to express thanks to Senior Inspector Sanford Shoults, Inspector Ralph Baker, and Lieutenant Francis Davey for their contribution to this research. They were responsible for maintaining the records used and were helpful in making the basic tabulations possible.

Table 1
Homes in Which Excessive Drinking Led to Abuse of Boy

Excessive drinking	Total	Number Repeaters	Non-repeaters	Contribution to chi square
Present	7	3	4	4.67
Not present	197	40	157	0.03
Not stated	3	0	3	1.50
Total	207	43	164	6.20

$\chi^2 = 6.60$; $df = 2$; $0.05 > p > 0.02$

Table 2
Attitude of Parents toward Police Officer

Attitude	Total	Number Repeaters	Non-repeaters	Contribution to chi square
Cooperative	190	34	156	0.81
All others	17	9	8	8.17
Total	207	43	164	8.99

$\chi^2 = 8.99$; $df = 1$; $p < 0.01$

Two of the significant tables related to school. As shown in Table 3, almost half of the boys who, at ten, were receiving failing grades in school continued to get in trouble. Also, as shown in Table 4, a similar proportion of boys who were in special classes (in Detroit this means classes for mentally retarded youngsters) became repeaters. Here again, however, the statistical significance of the results is clouded by the small numbers involved.

Table 3
School Marks

School marks	Total	Number Repeaters	Non-repeaters	Contribution to chi square
Good	27	3	24	1.83
Average	119	24	95	0.05
Poor	45	9	36	0.00
Failing	16	7	9	6.56
Total	207	43	164	8.44

$\chi^2 = 8.44$; $df = 3$; $0.05 > p > 0.02$

Table 4
Grade in School

Grade	Total	Number Repeaters	Non-repeaters	Contribution to chi square
3rd or lower	41	9	32	0.16
4th	69	14	55	0.00
5th or higher	82	13	69	1.20
Special class	15	7	8	6.67
Total	207	43	164	8.03

$\chi^2 = 8.03$; $df = 3$; $0.05 > p > 0.02$

The statistical fault does not pertain to the results reported in Table 5. A high proportion of the repeaters had two or more brothers. The statistical significance is clear. Interestingly, there was no such relationship with number of sisters. Another index to living conditions, the type of dwelling unit occupied by the boy's family, as shown in Table 6, was also related to repeating. Those boys who lived in apartments or rooming houses were prone to become repeaters. The chance expectations in the key cells of this table were amply large to warrant a rejection of the null hypothesis without cavil. A third item related to family economic conditions was the need for a boy to employ part of his funds for clothes, maintenance, or similar uses. Among ten-year-olds this is rare, as indicated in Table 7. It was more prevalent for those boys who later became repeaters, but the numbers concerned were too small to warrant accepting the chi square at its face value.

The last item to be reported has a strange history. The present study is one in a series begun at the request of police officials. Throughout there has been much give and take on topics to be investi-

Table 5
Number of Brothers

Number of brothers	Total	Number Repeaters	Non-repeaters	Contribution to chi square
0	61	10	51	0.88
1	66	6	60	5.80
2	39	14	25	5.66
3 or more	41	13	28	3.89
Total	207	43	164	16.23

$\chi^2 = 16.23$; $df = 3$; $p < 0.01$

Table 6
Type of Dwelling Unit

Type of unit	Total	Number Repeaters	Non-repeaters	Contribution to chi square
Single-family house	93	22	71	0.59
Flat	81	10	71	3.65
Apartment or rooming house	30	11	19	5.21
Not stated	3	0	3	1.50
Total	207	43	164	10.95

$\chi^2 = 10.95$; $df = 3$; $0.02 > p > 0.01$

Table 7
Boy's Disposition of Money

Disposition	Total	Number Repeaters	Number Non-repeaters	Contribution to chi square
All spent for entertainment	193	37	156	0.31
All or part required for clothes, maintenance, etc.	14	6	8	3.82
Total	207	43	164	4.13

$\chi^2 = 4.13$; $df = 1$; $0.05 > p > 0.02$

gated. When the "history sheets" from which our data are taken were designed, the police seemed anxious to include a judgment as to the "type" of a boy. One of the "types" they listed was "Peck's Bad Boy," by which they meant a youngster who frequently did things which angered the adults of the neighborhood. Although hardly the sort of "type" that had meaning to psychologists, it was included as a token of our willingness to accept police estimates of what was important. It must be confessed that it would be more accurate to say the item was kept in to humor the police. The fact is that in all of the studies we have made, this item turns out to be highly predictive. Whatever it is in a boy that makes an experienced policeman check him as a "Peck's Bad Boy," it involves recognition of behavior trends which do indicate an important and persistent propensity toward delinquency. As shown in Table 8 this item once again proved predictive. The numbers involved were sufficiently large to warrant rejecting the null hypothesis at the 1 per cent level of confidence without reservation. Remember in this case the judgment was recorded at the time of first contact and proved predictive of conduct two years later.

Table 8
"Type" of Boy

"Type"	Total	Number Repeaters	Number Non-repeaters	Contribution to chi square
Normal	143	21	122	3.42
"Peck's Bad Boy"	52	20	32	9.34
All others	12	2	10	0.44
Total	207	43	164	13.20

$\chi^2 = 13.20$; $df = 2$; $p < 0.01$

Although detailed examination of the tables did bring out some predictive items, these did not fall into the pattern required to support our second hypothesis. The findings in relation to home and school conditions were inconclusive. Clear-cut results pointed to frequent migration (as implied by living in rooming houses), the presence of many brothers, and the existence of strong adult-annoying characteristics as differentiating the repeaters from the nonrepeaters.

Discussion

The major finding of this study is that only a small proportion of ten-year-olds against whom complaints are registered with the police continue to get in trouble. Before accepting this as bearing on the point that preadolescent misconduct may be symptomatic of conditions which are self-correcting or which disappear as a result of maturing, we must weigh the possibility that many of these boys lacked police contacts because they had moved from Detroit or that they may have received therapy.

Three facts would seem to indicate that migration from Detroit was not a significant factor. First, the rooming house occupants, the group most likely to be on the move, were more heavily represented among the repeaters than the nonrepeaters. Second, the period, 1948 to 1950, was a relatively stable one during which there was little emigration from the city. Third, the fact that there were statistically significant differences not related to economic circumstance differentiating between repeaters and nonrepeaters would nullify any contention that the two groups were alike except that the nonrepeaters happened to leave the police jurisdiction. We can safely discount the argument that the ten-year-olds had left town in any major proportion.

As to the possibility that they had received therapy, here we cannot be as confident. The indications that this was not the case are indirect. As previously mentioned, referral to the court clinic in 1948 tended to be reserved for unusual cases, especially those having sex offenses. Our data did include the offense with which the boys were charged. On this count there was no significant difference beween the repeaters and the nonrepeaters. However, there is no way of knowing exactly what each family did as a consequence of a boy getting in trouble. We do know that during this period a special arrangement existed whereby the police could get speedy action on any referrals to social agencies and that this machinery was used for only four cases. As far as the police records show, there was almost no use made of the social and clinical resources of Detroit for the boys with whom we are concerned. In general, these services have been brought into the picture for the older youngsters who have longer records and whose families have given up hope of doing anything on their own. For the 20 per cent of the ten-year-olds who have become repeaters, this is most unfortunate.

Summary

A follow-up study was made of 207 boys who as ten-year-olds in 1948 had been the subjects of complaints to the Detroit police. Only 43 of this group had been in trouble with the police during 1950. The items recorded at the time of the first contact which proved predictive of repeating were (*a*) having two or more brothers, (*b*) living in an apartment or rooming house, and (*c*) having a reputation as a "Peck's Bad Boy."

Received December 30, 1952.

References

1. Blair, A. W., & Burton, W. H. *Growth and development of the preadolescent.* New York: Appleton-Century-Crofts, 1951.
2. Redl, F. Preadolescents—What makes them tick? *Child Study,* 1944, **21**, 44-48.
3. Wattenberg, W. W. Factors associated with repeating among preadolescent "delinquents." *J. genet. Psychol.,* in press.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

A Note on the K Scale of the Minnesota Multiphasic Personality Inventory

Anders Sweetland and Herbert Quay
Florida State University

The Minnesota Multiphasic Personality Inventory (MMPI), designed by Hathaway and McKinley [3], is a 560-item inventory comprised of several scales designed to measure in the areas of: hypochondriasis (*Hs*), depression (*D*), hysteria (*Hy*), psychopathic deviate (*Pd*), masculinity-femininity (*Mf*), psychasthenia (*Pt*), paranoia (*Pa*), schizophrenia (*Sc*), and hypomania (*Ma*). It also has four "validity" scales, designed to measure test-taking attitudes. Among these is the *K* scale [5]: a measure of "plus-getting" (a tendency to get poor MMPI scores) and "defensiveness" (a tendency to get good scores).

Despite the recommendations of several investigators [1, 7, 9], most clinicians will still venture profile diagnosis of the MMPI. Although the evidence indicates that the profile should be used mainly to differentiate between the neurotic, the psychotic, and the normal, we hazard that few psychologists will be able to avoid searching the profile for clues about a client. Since profile analysis seems to be inevitable, it seems axiomatic that it should be as adequate as possible.

Clinicians tend to approach the MMPI profile by one of two avenues: (*a*) as a nosological aid, or (*b*) as a projective technique. The latter generally begins with a careful consideration of the test items. With these well in mind the various scores become a frame of reference rather than a measure of hypochondriasis or hysteria. It is presumed that the empathetic counselor, knowing the items well, can "slip on the garb of the profile" and view the world through the eyes of the client. This, patently, is not rigorous psychology, and few pretend that it is. On the other hand, profile analysis of this kind can be uncannily accurate.

Profile analysis of this second type is not as esoteric as it sounds. The clinician thoroughly familiarizes himself with items in the various subscales. A high score tells him that a particular set of items was answered in a given manner. A low score indicates the opposite. The job of learning the items is facilitated by the fact that most of them are about what is expected. There are exceptions, but in doubtful cases one can refer to the aberrantly scored items.

The *K* scale, however, does not lend itself readily to this practice. The items fall into no apparent pattern. If they suggest anything, they seem to deny having pathological symptoms. As a consequence, the clinician uses the scale for what it was designed: as a control of test-taking attitude.

The possibility that the *K* scale might measure something other than test-taking attitude was suggested by its fluctuation during an experiment with hypnotic dreams [8]. The results indicated that symbolized dreams were produced by well-adjusted people and not by maladjusted people. The measures of adjustment were the MMPI and the Social Security-Insecurity Test (SI) of Maslow, Hirsh, Stein, and Honigman [4]. The SI is a 75-item inventory designed to measure in the area of social security.

The correlations between the adjustment scales of the MMPI and the number of symbolized dreams were low (Circa —.30) but all were negative. In contrast the correlation between *K* scores and symbolization was high and positive (.72) as was the correlation between the SI and symbolization (.57). The situation was that *K* correlated high with dream symbolization and the latter was produced only by well-adjusted people.

The dream experiment was repeated under slightly different conditions. The statistical results were essentially the same: high correlation between *K* and SI scores when paired with symbolization and negative correlations between symbolization and the adjustment scales of the MMPI. The relationships were evidently not due to chance behavior of the data. Apparently *K* was a measure of adjustment.

As a check, the *K* scale and the SI were given to seventy-five college students. The resulting correlation was 0.68. A second sample of ninety-one college students yielded a correlation of 0.74.[1] It would seem from this that the *K* scale measures in much the same area as the SI.

Two places in the literature suggest similar "misbehavior" of the *K* scores. The first was in the work of Gough [2], whose data suggest that when faking is attempted, *K* scores are depressed in a marked fashion although *K* scores of psychiatric groups ordinarily are slightly below average. (This was inferred from his data. Faked profiles show *F—K* scores of 14 to 41 while bona fide scores from normal and psychiatric groups range from—3 to —9.)

Gough's data, conforming with that offered originally by Meehl and Hathaway [5], also indicate that college students are characterized by elevated *K* scores. His average *K* is approximately 18 for this group (again inferred from his data), while *K* for average adults is approximately 13.

A factor analysis of the MMPI by Wheeler, Little, and Lehner [9] provides another opportunity to consider the *K* factor. Like Meehl and Hathaway they find that *K* before factoring correlates negatively with all the subscales except *Hy*. Similarly, they find a marked negative relationship when correlating *K* with *Pt* and *Sc*. (The range is from −.55 to −.67). The inference seems to be that the *K* scale measures something in common with the *Hy* scales, but something opposite the *Pt* and *Sc* scales.

In the factor analysis, *K* again appears to behave unusually. With the exception of the second factor extracted from the normal group, *K* never appears in the same factor as *Hy*, and it is the only scale that yields high negative loadings. Further, it tends to be significant only in the first factor extracted, a factor which has high positive loading of *Pt* and *Sc*.

Our interpretation differs somewhat from that of Wheeler *et al.* They interpret *K* as representing ego "defense." We believe that a more logical interpretation might be made by considering *K* as being one end of a continuum with *Pt* and *Sc* at the opposite end. This is based on the fact that when *K* and the other two scales appear in the same factor, the *K* loadings range from −.60 to −.70 while the *Pt* and *Sc* loadings are .90 and higher. From this we venture that what is measured by the *K* scale is the opposite of that measured by the *Pt* and *Sc* scales. We owe Wheeler, Little, and Lehner an apology for selecting this limited aspect of their excellent study. The interested reader is urged to consult the original article.

One further clue is provided. In a follow-up study [6], the hypnotic dreams were rated as being "extrapunitive," "intrapunitive," "impunitive," and "no feeling." The correlations between these and the *K* scores were: extrapunitive, −.81; intrapunitive, −.33; impunitive, —.20; and no-feeling, +.61. The no-feeling dreams were produced chiefly by the well-adjusted subjects. The results suggest that *K* tends to be opposite from extrapunitiveness and closely associated with good adjustment. Although the two highest correlations are reliable at the 1 per cent level, the interpretation is made with caution because of the small sample: dreams = 405, subjects = 16.

Summarizing these bits of information we might say that normal *K* scores seem to be associated with: (*a*) social security; (*b*) emotional adjustment; (*c*) a tendency to symbolize hypnotic dreams; (*d*) being a college student; (*e*) nonfaked profiles; (*f*) the opposite of what is measured by the *Pt* and *Sc* scales; (*g*) nonextrapunitiveness.

If we were to characterize the person described by this summary, he would appear to be a therapeutic ideal: a person who is comfortable in his social contacts, above average in intelligence, and little bothered with doubts and indecisions. He seems to have an ability to live each day as it comes and to "sleep well o' nights." The indication is that this is a core part of his personality and not a façade that he erects to enable him to meet the day.

---

[1] The correlation between the SI and the social IF scale of the MMPI was .77, the correlation between *K* and the MMPI SIF scale .48.

In view of the above, again emphasizing that the data are tenuous, it is our feeling that *K* measures not only the degree of "defensiveness" or "plus-getting" with which a person takes the MMPI, but that it may also be a measure of healthy emotional adjustment.

Summary

The possibility that the *K* scale of the MMPI might measure something in addition to test-taking attitude was discussed. The suggestion was made that K scores, when within normal ranges, may possibly be a measure of personality integration.

Received December 8, 1952.

References

1. Benton, A. L., & Probst, Kathryn A. A comparison of psychiatric ratings with Minnesota Multiphasic Personality Inventory scores. *J. abnorm. soc. Psychol.,* 1946, **41**, 75–78.
2. Gough, H. G. The *F* minus *K* dissimulation index for the Minnesota Multiphasic Personality Inventory. *J. consult. Psychol.,* 1950, **14**, 408-413.
3. Hathaway, S. R., & McKinley, J. C. *Manual for the Minnesota Multiphasic Personality Inventory.* New York: Psychological Corp., 1943.
4. Maslow, A. H., Hirsh, E., Stein, Marcella, & Honigman, Irma. A clinically derived test for measuring psychological security-insecurity. *J. genet. Psychol.,* 1945, **33**, 21-41.
5. Meehl, P. E., & Hathaway, S. R. The *K* factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *J. appl. Psychol.,* 1946, **30**, 525-564.
6. Quay, H. C. Emotions in hypnotic dreams. Unpublished master's thesis, Florida State Univer., 1952.
7. Schmidt, H. O. Test profiles as a diagnostic aid: The Minnesota Multiphasic Inventory. *J. appl. Psychol.,* 1945, **29**, 115-131.
8. Sweetland, A., & Quay, H. An experimental investigation of the hypnotic dream. *J. abnorm. soc. Psychol.,* 1952, **47**, 678–682.
9. Wheeler, W. M., Little K. B., & Lehner, G. F. The internal structure of the MMPI. *J. consult. Psychol.,* 1951, **15**, 134–141.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

The Archimedes Negative Aftereffect as an Indication of Memory Impairment

Lloyd S. Standlee
Institute of Educational Research, Indiana University

Visual fixation upon a slowly rotating Archimedes spiral produces the illusion of expansion or contraction. Which illusion is experienced depends upon the direction of rotation. The opposite illusion (negative aftereffect) is experienced when the gaze is transferred to any other fixed object. With clinical judgment as the criterion of memory impairment, Freeman and Josey report " . . . a strong probability of correlation between the negative aftereffect of the Plateau [Archimedes] spiral and the impairment of memory" [1, p. 797].

The present experiment utilized a more objective criterion to further test the possibility of this relationship between a visual phenomenon and memory ability. Twenty-five "psychotics" and 16 "normals" were tested on ability to experience the negative aftereffect and on the Wechsler Memory Scale [2]. The psychotics were tested again approximately eight hours after the administration of electric shock.

All of the 25 psychotics experienced the negative aftereffect in the first testing session. Twenty-three experienced it for both the expanding and the contracting illusion in the second testing session. One experienced it for the expanding illusion only; one for the contracting illusion only. All of the 16 normals experienced the negative aftereffect for both the expanding and the contracting illusion.

The psychotics averaged 71 MQ before shock and 75 MQ after shock. The gain of 4 MQ points is significant at the 1% level of confidence (t test on individual MQ gains). The normals averaged 120 MQ points. Combining the two groups, the MQ's ranged from below 48 to above 143—the upper and lower limits of the Wechsler Memory Scale.

Thus, Freeman and Josey's "strong probability of correlation" between the visual phenomenon and memory ability is not confirmed. All but two subjects were perfect in their ability to experience the negative aftereffect. In memory ability, however, the subjects were distributed fairly evenly through the entire range of Memory Quotients on the Wechsler Memory Scale. In other words, memory impairment to the extent of below 48 MQ does not prevent a person from experiencing the negative aftereffect.

The second testing session was included to determine whether memory ability would be impaired following electric shock. If the Wechsler performance and ability to experience the negative aftereffect were to drop, an additional measure of memory impairment would be available. Such was not the case, however. The MQ's improved significantly (in the statistical sense), and the visual phenomenon was not affected. Though the improvement in MQ's was statistically significant, it was confounded with practice effect. Therefore, it cannot be concluded that memory ability was improved by electric shock. It can be said, however, that memory ability was not impaired.

It is concluded (*a*) that electric shock does not impair memory ability; (*b*) that electric shock does not impair ability to experience the Archimedes negative aftereffect; and (*c*) that there is no relationship between memory ability and ability to experience the Archimedes negative aftereffect.

Received April 6, 1953.
Early publication.

References

1. Freeman, E., & Josey, W. E. Quantitative visual index to memory impairment. *Arch. Neurol. Psychiat.*, 1949, **62**, 794–797.
2. Wechsler, D. A standardized memory scale for clinical use. *J. Psychol.*, 1945, **19**, 87–95.

NEW BOOKS AND TESTS

Books

Baker, Harry J. *Introduction to exceptional children.* (Rev. Ed.) New York: Macmillan, 1953. Pp. xvi + 500. $5.00.

A college textbook planned for the training of school personnel. Exceptional children are often a complicating problem to the smooth running of the classroom. Baker deals effectively with the problems of unusual or handicapped children and seeks to give a teacher an awareness of their educational needs. The present edition is a revision of the original publication of 1944. There have been only a few minor changes in content, but the bibliography and index have been revised and brought up to date. — *B. M. L.*

Blum, Gerald S. *Psychoanalytic theories of personality.* New York: McGraw-Hill, 1953. Pp. xviii + 219. $3.75.

In so modestly brief a volume, it is something of a surprise to find such a comprehensive survey of psychoanalysis. All shades of analytically-oriented theory are represented—Freud and the orthodox—the early deviants, Adler, Jung, and Rank—the neo-analysts, including Fromm, Horney, Sullivan, and others. The text proper, organized in terms of developmental stages, is a clear and unbiased statement of the analysts' theories. All are represented fairly, and the author does not impose his preferences. In his "notes," which occupy a third of the volume, Blum permits himself the luxury of criticism and opinion, and brings to bear much of the relevant evidence from research. He ably supports his thesis that psychoanalytic views provide the most promising present basis for a theory of personality, but that these views need to be sorted out, evaluated by research, and progressively refined. The book is a valuable reference for all interested in psychoanalysis—as who is not, these days—but especially for those disposed to approach it with a research orientation. —*L. F. S.*

Brower, Daniel, & Abt, Lawrence E. (Eds.) *Progress in clinical psychology.* Vol. 1, Secs. 1 & 2. New York: Grune & Stratton, 1952. Pp. xi, xxiii + 564. 2 vols. $5.75, $5.00.

After a cogent introductory chapter by Abt, the rest of this forty-two chapter, two-volume work is contributed by 38 different authors, most of them clinical psychologists. In the first volume the aim has been "to provide as complete a coverage as possible for the past six years in clinical psychology and to point up, in the process, as many stimuli as possible to further thinking and research." The writing is somewhat uneven, but the literature on the well-known intelligence and projective tests is well, if not exhaustively, covered. The following section on psychotherapy includes separate chapters on client-centered counseling, psychoanalytic theory and practice, play techniques, spontaneous art in therapy, and neurosis as a learning phenomenon. Volume 2 includes a section on developmental processes from infancy to old age, one on applications of clinical psychology to special areas (industry, rehabilitation, crime, etc.), and a section on approaches to clinical psychology which includes chapters on psychosurgery, cultural anthropology, social psychology, statistical methods, and P-technique. The volumes conclude with a thoughtful discussion of clinical psychology as a profession by Victor Raimy. The authors state their intention to issue follow-up volumes, similarly designed, every two or three years. This will necessarily cover much of the material surveyed yearly, but inevitably much more briefly, by the *Annual Reviews.*—*A. R.*

Cabot, Hugh, & Kahl, Joseph A. *Human relations: concepts and cases in concrete social science.* Cambridge: Harvard Univer. Press, 1953. Pp. xxi + 333, viii + 273. 2 vols. $4.75, $4.25.

Human Relations is the outgrowth of an undergraduate course in social science taught by the authors at Harvard. Volume I supposedly presents the concepts to enrich the cases presented in Volume II and to enable the student or reader to deal with the cases, preferably, one gathers, with the aid of an instructor. If Volume I were called "Readings in the Social Sciences," rather than "Human Relations," it would be more appropriately labeled. Per-

---

Note: The reviews were prepared by the Editor and the Associate Editors, who may be identified by their initials.

haps 90 per cent of the first volume consists of extensive excerpts, frequently complete chapters, from the works of well-known sociologists, psychologists, psychiatrists, philosophers, logicians, semanticists, and anthropologists. The authors usually supply brief introductions to these extensive quotations. The approach purports to be clinical, but it is anything but that. This volume is an excellent collection of first-rate writing in a variety of social science fields, but does not add up to a satisfactory treatment of the subject of human relations. Volume II consists of 33 cases portraying a great variety of situations in which "solutions" or "treatment programs" are not offered. This makes them useful for class discussion. The treatment of the cases is "common sense" rather than clinical. This is not said disparagingly since, to many, such an approach to human relations is preferred. It is questioned in this publication only because the source material presented in Volume I is largely from profound scholars and social scientists. The incongruity is striking, since the two volumes, which are intended to be used together, are practically unrelated.—*M. K.*

Hamrin, S. A. *Initiating and administering guidance services.* Bloomington, Ill.: McKnight & McKnight, 1953. Pp. xi + 220. $3.00.

The author states that his object is to aid and promote understanding and encouragement of guidance activities by school administrators. The topics discussed include the initiation of a program, in-service education, and public relations. In addition, chapters are devoted to descriptions of the traditional services such as orientation, individual study, counseling, occupational information, group activity, placement, and follow-up. There are many excerpts from unprinted reports on existing programs by school practitioners. It contains little content of a psychological nature. Although the author points out how few guidance workers have formal training, there is not enough in the volume to indicate that the best guidance means more understanding of human motivation, frustration, and conflict resolution both in client and counselor. Such an emphasis seems, to the reviewer, appropriate even in a book emphasizing administration.—*F. McK.*

Hoyles, J. Arthur. *The treatment of the young delinquent.* New York: Philosophical Library, 1952. Pp. xi + 274. $4.75.

A British clergyman undertakes "a careful examination of the lessons of history and the findings of modern science in the light of Christian principles." It leads him to the conclusion that "there is an alternative to punishment."—*E. L. K.*

Patterson, C. H. *The Wechsler-Bellevue Scales: A guide for counselors.* Springfield, Ill.: Charles C. Thomas, 1953. Pp. xiii + 146. $3.75.

This monograph is a thorough and competent review of the research literature about the Wechsler-Bellevue, with a bibliography of 250 entries. The presentation is well organized, and each major section ends with an integrative summary of the conclusions justified by the evidence. It is indeed a "guide for counselors" in the sense of displaying the values of the test and also its limitations.—*L. F. S.*

Trenaman, Joseph. *Out of step.* New York: Philosophical Library, 1952. Pp. xx + 223. $4.75.

A fascinating account of an attempt to rehabilitate young but seriously delinquent soldiers, carried out by the British army during the war period. A sample of 200 cases was selected for intensive study and compared with nondelinquent soldiers of the same age with respect to a large number of physical, psychological, and sociological variables. Although only one-third of the cases had a history of civilian offenses, the great majority appeared to have deep-seated problems likely to lead to conflict with either civilian or service authority. Most of the book is concerned with an analysis of the differences between delinquent and nondelinquent soldiers, but it includes many details of the remedial-therapeutic program, of four- to six-months duration, after which slightly over half of the cases made a satisfactory adjustment to army life.—*E. L. K.*

Weinland, James D., & Gross, Margaret V. *Personnel interviewing.* New York: Ronald Press, 1952. Pp. vii + 416. $6.00.

A general text dealing with interviewing as practiced in business and industry. It is apparently intended for students with little or no previous training in psychology, since many of the chapters read like an elementary text in psychology. For example, one chapter is devoted to individual differences and another to a discussion of psychological tests. Other chapters, however, discuss almost every technique of interviewing, and specific types of interviews used in the industrial setting. The only bibliographical references are those included in an occasional footnote.—*E. L. K.*

Tests

Grassi, Joseph R. *The Grassi Block Substitution Test for Measuring Organic Brain Pathology.* Manual, pp. 75 ($3.00). Springfield, Ill.: Charles C Thomas, 1953.

The *Grassi Block Test* is an ingenious modification of the Kohs technique which involves two novelties, the use of three-dimensional reproduction, and the observation of ability to "shift" to a new color pattern. Each of five models, constructed from Kohs's blocks, is presented as stimulus material for four progressive responses: (*a*) copying the top design, (*b*) copying the top design "with different colors," (*c*) copying the three-dimensional design, and (*d*) copying the three-dimensional design with a shift of colors. In view of the well-established findings that organic cases have difficulty with complex spatial perceptions, and also with ability to "shift," this double-barreled testing method rests on a sound rationale. Both an objective scoring system and qualitative indications are used to interpret the test. The retest reliability is reported as .85, and validity data on 276 cases show a clear separation of organics from normals and undeteriorated schizophrenics with little overlapping. Although the test needs further cross validation, it seems likely to become an interesting and efficient method for studying the psychological effects of brain pathology.—*L. F. S.*

Kuhlmann, F., & Anderson, Rose G. *Kuhlmann-Anderson Intelligence Tests.* (6th Ed.) 9 levels, ages 5-adult. 1 form at each level. 20-30 (40-45) min. Test booklet, each level ($2.40 per 25), with key, directions, pp. 42, and handbook, pp. 40; master manual, pp. 151 ($2.50); specimen set, 1 level ($1.00). Princeton, N. J. (188 Nassau St.): Personnel Press, 1952.

It is gratifying to see a useful psychological instrument kept up to date by continued research. Although the test materials are basically the same as those first published in 1927, new analyses of reliability, discriminative power, and norms have been made in 1951. Among group tests, the Kuhlmann-Anderson has always been distinguished by a clinical or "individual" flavor. It is not the easiest of such instruments to administer and score, but qualified users have felt rewarded for their efforts. Either timed or untimed administration is possible, with reliabilities centering around .95. The manual provides recently verified norms, and extensive statistical data including subtest intercorrelations. The new edition gives the impression that the test is by no means outdated, but is worthy of more years of service.—*L. F. S.*

Books Received

Allen, Robert M. *Introduction to the Rorschach technique.* New York: International Universities Press, 1953. Pp. 126. $3.00.

Crawford, Claude C., Cooley, Ethel G., Trillingham, C. C., & Stoops, Emery. *Living your life.* Boston: Heath, 1953. Pp. xv + 448. $2.80.

Devereux, George. (Ed.) *Psychoanalysis and the occult.* New York: International Universities Press, 1953. Pp. iv + 220. $3.00.

Field, Minna. *Patients are people.* New York: Columbia Univer. Press, 1953. Pp. viii + 244. $3.75.

Natenberg, Maurice. *Freudian psycho-antics.* Chicago: Regent House, 1953. Pp. 101. $2.00.

Piaget, Jean. *The origins of intelligence in children.* New York: International Universities Press, 1953. Pp. xi + 419. $6.00.

Podolsky, Edward. (Ed.) *Encyclopedia of aberrations.* New York: Philosophical Library, 1953. Pp. viii + 550. $10.00.

Reid, Joseph H., and Hagan, Helen R. *Residential treatment of emotionally disturbed children.* New York: Child Welfare League of America, Inc., 1952. Pp. ix + 313. $3.50.

Slavson, S. R. *Child psychotherapy.* New York: Columbia Univer. Press, 1952. Pp. xiii + 332. $4.50.

Tuckman, Jacob, and Lorge, Irving. *Retirement and the industrial worker.* New York: Bureau of Publications, Teachers College, Columbia Univer., 1953. Pp. xvi + 105. $2.75.

Journal of Consulting Psychology
Vol. 17, No. 5, 1953

A Scale for Rating the Responses of the Psychotherapist

Rex Madison Collier
University of Illinois[1]

The concept of applicable continua for quantitatively describing the varying verbalizations of the psychotherapist has previously been presented [3]. The purpose of this article is to describe some initial attempts to quantify one of these continua, and to outline some of the applications that seem to be feasible from this approach. In the article referred to above two continua were selected for special attention from among those suggested. These were (*a*) the uncovering or interpretive continuum and (*b*) the responsibility-balance continuum. In the present paper attention will be given only to the first of these.

With reference to the interpretive continuum, it is noted that, among other things, the therapist's verbalization is designed to aid the client in arriving at a clearer awareness than has existed before of some feeling, attitude, relationship, or aspect of his personality. Opportunities for changes in meaning are assumed to be facilitated. Examination of verbatim transcriptions of psychotherapy interviews reveals variation, not only among therapists, but for the same therapist from one response to another, in the degree to which a therapist's responses seem designed to facilitate an increased awareness of a relatively unrecognized aspect of a client's tendencies. It is neither denied nor forgotten that therapists do other things as well. They "structure relationships"; they indicate an "accepting," "permissive," "understanding" approach; at times, some may include "reassurance," "advice," and "supportive" therapy; they may also ask questions. Obviously, no particular school of therapy is here being emphasized. It is not the purpose of this series of studies to carry the aegis for any system to the exclusion of others. The initial problem of this study was to devise classificatory descriptions of a larger number of steps along the uncovering-interpretive continuum than was done in the former article. In this earlier publication five steps were indicated. It remained to determine (*a*) if additional steps could be described, (*b*) if judges with experience in psychotherapy could reliably distinguish differences among these positions or along such a scale, and (*c*) to what extent would the original assumptions about relative scale positions be borne out by ratings of experienced judges.

The descriptive items that appear below together with the accompanying instructions were submitted to 40 psychotherapists—analytic psychiatrists and clinical psychologists. Twenty-six returned the material with completed ratings. Several of the remaining 14 explained that the approach seemed too different from their theoretical orientation to provide a basis for making the requested judgments. One thought this to be an attempt to stereotype therapy and declined on that basis; another misunderstood the directions, assuming that the experimenter wished him to score himself in retrospect as he had performed therapy. The wide variety of approaches to the therapeutic relationship was inescapably evident from the letters of those who declined to participate. The variety of theoretical backgrounds for psychotherapy increases the difficulty of adequate communication. This problem is illustrated by the critic who felt that the writer was attempting to stereotype psychotherapy, whereas one of the main interests behind this series of studies is to test the possibility of liberating psychotherapy from a stereotyped procedure.

---

[1] This article was completed after the writer accepted a position as Chief Psychologist at the Jefferson Barracks Veterans Administration Hospital.

The rating sheet was entitled "A Study in Discrimination" and the task was introduced by the following instructions:

> You are being asked in this study to judge or rate types or kinds of responses which psychotherapists make in helping a patient work through his problems. It may be assumed for purposes of this study that therapists' responses are designed, among other things, to aid the patient in developing insights, penetrating resistances, or uncovering hitherto unrecognized aspects of the personality. The degree to which the therapist response is designed to penetrate typical resistances or uncover, etc. may vary from little or nothing to very much. Furthermore, it must be emphasized that no implications are indicated in regard to what are appropriate or desirable therapist responses. Such judgments are made in terms of other criteria. Since this is a different type of evaluation or judgment, your discriminations in terms of the present variable are to be made independently of theoretical convictions you may have about what a therapist should do.
>
> Below are descriptions of these types of responses. Ratings must, of course, be in terms of general implications, since it is recognized that the context of a specific therapy situation would introduce some variation. After reading a description of a type of therapy technique, indicate by encircling a numerical position on the scale (1 to 11) the degree to which that kind of response seems in general designed to penetrate typical resistances, uncover or bring into clearer recognition an aspect of personality. Read all descriptions before rating any of them. Remember 1 = low degree of the variable described; 11 = high degree.

The descriptive items are presented below. Except for item *a,* the scale which appeared after each item has been omitted.

a. The therapist repeats the statement the patient has just made.

1 2 3 4 5 6 7 8 9 10 11

---

b. The therapist suggests a relationship between what the patient has just said and what he said at an earlier time.

c. The therapist responds by a nonverbal vocalization, e.g., "M-hm."

d. The therapist abstracts common dynamic elements or trends from repetitive occurrences of patterns of feelings, thoughts, and action when such patterns are dominated by a dissociated, rejected, or repressed experience or trend and brings this relationship to the recognition of the patient by any of a variety of interpretive procedures.

e. The therapist selects a portion of the patient's response and repeats it.

f. The therapist remains silent.

g. The therapist condenses or epitomizes what the patient has just said in terms of what seems to be the essential immediate meaning of the patient's response.

h. The therapist suggests a relationship between current attitudes and reactions of patient and early experiences or tendencies when such relationships have thus far gone unrecognized or only symbolically represented.

i. The therapist suggests a relationship between a current patient response and some early experience or tendency when such relationship has already been vaguely recognized by the patient.

j. The therapist phrases his response so as to indicate a recognition of the feeling or emotional aspect of what the patient has just said.

k. The therapist phrases his response so as to indicate a recognition of the content of what the patient has just said.

Results and Interpretation

The data collected in the manner described above were treated according to Thurstone's method (10) for attitude-scale building. Percentile points were determined graphically. Table 1 shows the results, based on an N of 26. The entries of the table are arranged in descending order of the Q_2 values.

Table 1

Initial Scale Values of the Uncovering Continuum

Descriptive Items	Quartile Points Q_1	Q_2	Q_3	Q_3-Q_1
d	9.4	10.3	10.6	1.2
h	7.8	9.1	9.7	1.9
i	7.4	8.2	9.0	1.6
b	4.8	6.5	7.6	2.8
j	4.5	5.5	6.5	2.0
g	3.4	5.4	6.6	3.2
e	2.0	3.5	4.7	2.7
k	1.8	2.7	3.7	1.9
a	0.6	1.5	2.8	2.2
c	0.6	1.4	2.1	1.5
f	0.3	0.7	2.4	2.1

From Table 1 it is apparent that these descriptive categories of therapists' responses represent discriminable differences. The scale distances from one item to the next, however, are not uniform, nor was it expected that they would be. It is theoretically possible, however, to construct sufficient additional descriptive categories so that a later more ambitious scaling attempt could provide a residuum of items which would represent a more uniform sequence of steps from one scale extreme to the other. Items used here tend to represent about five regions or levels according to their tendency to cluster. Since the average semi-interquartile range is 1.05, it is assumed somewhat arbitrarily that scale distances greater than 1 between items begin to represent reliable differences. Thus, near the zero end there are three items within a range of 1.1; these are *f, c,* and *a.* Silence, *m-hm,* and repetition appear at scale positions of 0.7, 1.4, and 1.5, respectively. It may be a mistake to put silence in this group for theoretical reasons and it does seem to be somewhat by itself. Clearly, however, *m-hm* and repetition are scaled in very close proximity. The second level is represented by *k* and *e,* wherein simple recognition of content and repetition of a selected portion of the patient's response seem more similar to each other than to items on either side.

In the midpart of the scale three items occur within the range of 1.1, while the cluster is separated on either side by nearly 2 scale points. These items, *g, j,* and *b,* are categories of therapists' responses which seem obviously designed to provide more aid in the reassociative process than those preceding. That the scaled apparent intent of the therapist corresponds with the implied result, i.e., that the reassociative process is facilitated, is a hypothesis which remains to be tested. However, it would seem more penetrating or revealing, for example, to recognize feeling ($j = 5.5$) than simply to recognize content ($k = 2.7$); or, to suggest a relationship between current and previous patient verbalization ($b = 6.5$) would seem more revealing than simply to repeat a portion of a patient's response ($e = 3.5$).

The writer recognizes, of course, that specific therapy situations under which these categories of responses occur can cause considerable variation in a therapist's intent and in the associative facilitation for the patient. Because of this, one may be somewhat surprised that the *Q* measure of reliability does not indicate more variability than it does.

The fourth level is represented by *i* and *h* at 8.2 and 9.1, respectively. Both of these items describe therapists' responses as suggesting relationships. This also occurred for *b* at 6.5; but at 8.2 and 9.1 the relationship dealt with has gone beyond two temporally separated patient responses and involves relationships either only vaguely recognized or symbolically represented by the patient. Again, it would seem obvious that the therapist is using a more penetrating technique than at levels scaled at lower values.

The fifth level is represented in this study by *d,* scaled at 10.3. Here the therapist is again suggesting relationships, but now it is on a broader basis. He attempts to help his patient see common dynamic elements running through a variety of experiences and activities when these elements have tended to elude the patient because of dissociative emotional resistances. Perhaps it should be said again that the purpose of this paper is concerned neither with the "whether" nor the "when" problems of appropriateness of these categories of therapist activity. Some therapists would be appalled by the idea of their ever using a response scaled higher than the one at 5.5. Others do not hesitate to use a variety of responses ranging over the whole scale. Only extensive evaluative studies can eventually provide the answers of "if" and "when" these categories are appropriate.

Pilot-study applications of this scale have been made to such material as found in Snyder's *Casebook* [9], Axline, *Play Therapy* [1], Rogers' case of Herbert Bryan in *Counseling and Psychotherapy* [7], Berg's *Deep Analysis* [2], Deutsch's *Applied Psychoanalysis* [4], and Eidelberg's *Take Off Your Mask* [5]. In the Axline material the writer has found nothing to rate beyond a scale value of four; in Snyder and Rogers the upper scale value limit is six. The therapist responses in Berg, Eidelberg, and Deutsch seem to range over the whole scale. On well-known theoretical grounds, the "nondirective" group imposes limitations on themselves which the analysts do not. In terms of therapy goals, e.g., quality and rates of improvement, it remains to be determined what limits are appropriate for age, sex, type of problem, stage of therapy, etc.

In general, the hypothesis presented in the earlier publication [3] that there were discriminable differences along an uncovering or revelatory interpretive continuum has had initial confirmation. Furthermore, the steps along the scale as described at that time have occupied the sequence anticipated. However, the progressive increase in scale distance through the originally described five steps has not been confirmed in this study. In part this may be due to the accidents of illustrative selections, but it is apparently more assignable to larger discernible differences among the categories of responses at the lower end of the continuum than were hypothetically anticipated. Thus, there may be, for example, a greater scalable difference between repeating the client's statement and repeating a selected portion of the client's response than was originally assumed. Certainly one finds an analyst like Deutsch [4] using the technique of repeating a selected portion of a patient's response in a very penetrating manner.

In a second part of this study, 50 examples of verbal exchanges were submitted for rating to staff members and advanced graduate students who had had experience and training in psychotherapy. An eleven-point scale and similar instructions were used. Included in these selections were the examples in the initial article [3]. These selections were included in order to test the following hypotheses: (*a*) that these examples represent a sequence of discriminable differences along this continuum, (*b*) that the scale differences become progressively greater from one example to the next. The examples were as follows:

a. Pt. I hate my father, I hate my brother!

Th. You hate your father, you hate your brother.

b. Pt. I'll be ten years old. I'm in the fourth grade. My dad wants me to be in the fifth grade, but I tell him I don't want to be in the fifth grade. I like to fail.

Th. You like to fail.

c. Pt. The only thing is that I can't help but feel that I work against myself. I mean it's just so obvious to me that I do work against myself. And it seems rather silly to be one minute trying to do something about it, and then it seems so certain that there's something present that doesn't want to do anything about it. (Pause) It's such a contrary thing —the whole problem is just—

Th. So that the very time that one part of you is perhaps really taking hold of this whole situation, another part of you is just sabotaging the whole business.

d. Th. Do you remember that you had the same feeling in connection with mother, for whose health you were concerned, and that you behaved as if you were she? Do you remember that I reminded you of a statement of little Ralph who once said, "If I don't stop behaving so badly, it will kill her." Do you remember you said it?

Pt. That's right.

Th. And that you had to run home to see whether she were still alive and that you had a bad conscience as if you had done something wrong? Then you said something wrong means: Something is wrong with your heart or you did something wrong. Now, you said when your girl friend complained about her heart, you felt as if you were she.

e. Pt. He started to undress me and I embraced him. Suddenly, the image of my aunt appeared, and again I heard her warning. I grew frightened and ran away from Carl.

Th. What happened to the sexual excitement?

Pt. It disappeared as soon as I remembered my aunt's warning and grew frightened.

Th. Without this anxiety, you probably would not have heeded your aunt's warning, and you probably would have had sexual intercourse with Carl. From this point of view, the anxiety was a valuable help, because your sexual desire had destroyed your ability to resist your cousin. If your aunt's warning had appeared without anxiety, you would not have had the power to interrupt your sexual play and to escape unharmed. That was the situation at the time. Today, however, sexual intercourse is no longer a danger. Now you may conceive a child without being blamed for it. But you still behave as if you were a little girl. Why?

The scale values of these examples are given in Table 2, based on an N of 24.

Table 2

Scaled Examples of Psychotherapists' Responses

Examples	Quartile Points		
	Q_1	Q_2	Q_3
a	0.7	1.6	2.5
b	1.4	1.9	3.6
c	4.0	5.3	6.7
d	8.0	8.7	9.6
e	8.5	9.2	9.7

These data confirm the first hypothesis which stated that these verbal exchanges would represent discriminable differences along the continuum as described. The results with specific examples are similar to the results obtained when types of exchanges are rated. The finding implies that this kind of judgment can be made with a measurable degree of reliability, and that judges tend to perceive these differences in a similar way. The second hypothesis, which stated that there would tend to be progressive increases in scale distance from one example to the next, is only partly confirmed. The sequence of scale differences is as follows: $b-a=0.3$; $c-b=3.2$; $d-c=3.4$; and $e-d=0.5$. The first three differences become progressively larger, but the suggested pattern breaks down at $e-d$. These differences are, of course, in part, a function both of the method used and the examples selected. A paired-comparison method would yield somewhat different results. A longer scale might also make a difference, since some judges tend to be conservative in their use of extreme ends of a rating scale. Both the b–a and the e–d distances suggest this well known tendency of conservative raters to regress their ratings away from the extremes. That the selection of examples is important, also, is indicated by the fact that some in this set of 50 were consistently rated higher than the one used to illustrate the upper extreme in 1950. Such an example follows:

> Pt. My other dream seems to make no sense at all. I wanted my boy friend to kiss me and yet I felt I shouldn't hesitate to kiss him if I want to. Then as I started to kiss him, my nose got in the way; it seemed longer than usual, sort of like it had Pinocchio qualities. Jack noticed it right away. He seemed disappointed.
>
> Th. How did you feel?
>
> Pt. I believe I felt both surprise and somewhat elated.
>
> Th. Elated even though your boy friend was disappointed.
>
> Pt. That's what doesn't make sense. I don't want to lose him, but I seem to remember feeling that now I'm as good as he is.
>
> Th. Do you remember telling me that your father probably would have been happier if you had been a boy and that as you grew up you often competed with boys in boys' games? Now, it seems you would still like to compete; this time it's with your boy friend in playing the masculine role. You would not only like to be the aggressor but would also like to have a masculine genital organ.

The Q_2 value of this example is 10.2 with Q_3 and Q_1 at 10.6 and 9.1 respectively. The interquartile difference of 1.5 indicates little disagreement in the scale location of this item.

Applications of the Scale

Among the anticipated applications of a scale of this kind the following few are suggested:

1. Assuming that reliable criteria of patient improvement can be identified, we may relate the indices of improvement to scale levels used by the therapist.

2. Rogers and some of his students have insisted that insight occurs most frequently following simple acceptance by the therapist of what the client has said. Assuming that insightful responses may be identified and rated, we can relate the frequency of insight of varying degrees of comprehension to the scale level represented by the therapist's responses. According to Rogers' approach, we should expect the higher scale values of therapists' responses to be followed (*a*) by a reduced frequency of insight and (*b*) by an increased frequency of rejection by the client of these therapist responses. The writer's experience is such as to cast strong doubt on these theoretical positions.

3. Assuming that dependency and transference tendencies can be reliably identified in transcripts, we can relate these tendencies to scale value levels represented by the therapist. The nondirective school [8] would perhaps hypothesize that the more frequently the upper levels of the scale were used the greater would be the tendency for a dependency relationship on the therapist to develop.

4. Should therapy with psychotics and prepsychotics be supportive and noninterpretive? The literature on psychotherapy is in serious disagreement on the answer to this question. In the light of a scale approach such as the one above, the question might be revised as follows: "How interpretive should the therapist be in order to effect maximum improvement in the psychotic or prepsychotic patient?" Factors of age, sex, type of diagnosis, and duration of disturbance would have to be controlled in such an approach.

5. Fiedler's result [6] that experienced therapists are more alike regardless of school than are therapist and student within a school could be re-examined by this scaling approach. The writer's anticipation is that results from applying the interpretive continuum would not agree with Fiedler.

6. There are also applications in the training of students. One of these, certainly, is that this approach provides an integrative rationale which enables the student to feel secure with a more uncovering or interpretive approach to his patient under certain circumstances than if he is "all" analytic or "all" nondirective. Another application lies in the direction of testing the sensitivity of the student. Students' ratings of therapists' responses in examples of patient-therapist verbal interchanges seem to reveal varying degrees of student sensitivity to the implications of the interchange. Such ratings related to relative success in clinical courses, e.g., psychotherapy, might yield useful predictive results.

Summary

The continuum concept previously developed for quantitative application to the responses of the psychotherapist has been extended and initially applied. Types of therapist responses have been identified and scaled according to the Thurstone method of attitude-scale building. Reliability scores of scale positions indicate that judges with experience in psychotherapy tend to perceive the descriptive-scale items with a fair degree of consistency. The continuum concept developed previously for psychotherapy material has had initial substantiation. The hypothetical scale positions indicated in the previous article have been initially verified with minor exceptions. A variety of applications of this type of scale have been suggested for testing hypotheses concerning (*a*) relationships between patient improvement and quantitative variations of therapist activity, (*b*) relationships between type and frequency of insight and scale positions representing therapist responses, (*c*) relationships between content indications of transference and scale ratings of therapist responses, (*d*) relationships between scale ratings of therapists' responses and effective work with psychotics.

These are a few examples suggested only to illustrate what seem to be potentially fruitful applications of this continuum approach to some of our problems in psychotherapy.

Received January 5, 1953.

References

1. Axline, Virginia M. *Play therapy.* Boston: Houghton Mifflin, 1947.
2. Berg, C. *Deep analysis.* New York: Norton, 1947.
3. Collier, R. M. A basis for integration rather than fragmentation in psychotherapy. *J. consult. Psychol.*, 1950, **14**, 199–205.
4. Deutsch, F. *Applied psychoanalysis.* New York: Grune & Stratton, 1947.
5. Eidelberg, L. *Take off your mask.* New York: International Universities Press, 1948.
6. Fiedler, F. E. A comparison of therapeutic relationships in psychoanalytic, nondirective and Adlerian therapy. *J. consult. Psychol.*, 1950, **14**, 436–445.
7. Rogers, C. R. *Counseling and psychotherapy.* Boston: Houghton Mifflin, 1942.
8. Rogers, C. R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951.
9. Snyder, W. U. *Casebook of non-directive counseling.* Boston: Houghton Mifflin, 1947.
10. Thurstone, L. L., & Chave, E. J. *The measurement of attitude.* Chicago: Univer. of Chicago Press, 1929.

Journal of Consulting Psychology
Vol. 17, No. 5, 1953

An Ego-Strength Scale Which Predicts Response to Psychotherapy[1]

Frank Barron
Institute of Personality Assessment and Research
University of California

This paper reports the development and cross-validation of a scale which was originally designed to predict the response of psychoneurotic patients to psychotherapy. Consideration of the scale content and its correlates, however, suggests that a somewhat broader psychological interpretation be placed upon it, making it useful as an assessment device in any situation where some estimate of adaptability and personal resourcefulness is wanted. It appears to measure the various aspects of effective personal functioning which are usually subsumed under the term "ego-strength."

The scale consists of 68 items from the Minnesota Multiphasic Personality Inventory, selected from a total MMPI pool of 550 items on the basis of significant correlation with rated improvement in 33 psychoneurotic patients who had been treated for six months in a psychiatric clinic. The test responses of the patients were obtained before psychotherapy began, so that the scale, so far as logic of construction is concerned, is designed to predict whether or not after about six months of therapy the patient will have improved.

The sample of 33 patients was divided into two groups: 17 patients who were judged to have clearly improved, and 16 patients who were judged to be unimproved. Although the sample is small, the cases were intensively studied, and two skilled judges who had thoroughly acquainted themselves with the course of the therapy (although not themselves involved in it otherwise) were in considerable agreement (r of .91) in their independent ratings of degree of improvement. While one would not ordinarily base scale development on a sample of this size, it was reasoned here that a small number of well-studied cases who were classified with high reliability, and, as collateral evidence indicated [2], with high accuracy as well, would serve better than the practical alternative, which was to get a large sample in which the therapist's rating of outcome was accepted uncritically.

When the Improved and Unimproved groups were scored on this 68-item scale, the mean of the Improved group proved to be 52.7, that of the Unimproved group 29.1, a difference which is significant well beyond the .01 level (t of 10.3). The odd-even reliability of the scale in a clinic population of 126 patients is .76. Test-retest reliability after three months in a sample of 30 cases is .72.

The 68 items of the scale are presented below, arranged in groups according to the kinds of psychological homogeneities which, in the judgment of the writer, are involved in the item content.[2]

1. *Physical functioning and physiological stability.* 153. During the past few years I have been well most of the time (T). 51. I am in just as good physical health as most of my friends (T).

---

[1] This study was carried out for the most part at the Institute of Personality Assessment and Research, and was supported financially by a grant of the Rockefeller Foundation to that institute. Drs. Robert E. Harris and Timothy Leary supplied data for two of the cross-validating samples.

[2] The improved direction of response is given in parentheses after each item. The item numbering is taken from the booklet form of the MMPI. To save printing costs, a table giving the percentages of improved and unimproved patients answering each item "true" has been deposited with the American Documentation Institute. Order Document No. 4045 from ADI Auxiliary Publications, Photoduplication Service, c/o Library of Congress, Washington 25, D.C., remitting in advance $1.25 for 35-mm. microfilm or $1.25 for 6 by 8-in. photocopies.

174. I have never had a fainting spell (T). 189. I feel weak all over much of the time (F). 187. My hands have not become clumsy or awkward (T). 34. I have a cough most of the time (F). 2. I have a good appetite (T). 14. I have diarrhea once a month or more (F). 341. At times I hear so well it bothers me (F). 36. I seldom worry about my health (T).

2. *Psychasthenia and seclusiveness.* 384. I feel unable to tell anyone all about myself (F). 489. I feel sympathetic towards people who tend to hang on to their griefs and troubles (F). 236. I brood a great deal (F). 217. I frequently find myself worrying about something (F). 100. I have met problems so full of possibilities that I have been unable to make up my mind about them (F). 234. I get mad easily and then get over it soon (T). 270. When I leave home, I do not worry about whether the door is locked and the windows closed (T). 359. Sometimes some unimportant thought will run through my mind and bother me for days (F). 344. Often I cross the street in order not to meet someone I see (F). 241. I dream frequently about things that are best kept to myself (F).

3. *Attitudes towards religion.* 95. I go to church almost every week (T). 488. I pray several times every week (F). 483. Christ performed miracles such as changing water into wine. (F). 58. Everything is turning out just like the prophets of the Bible said it would (F). 420. I have had some very unusual religious experiences (F). 209. I believe my sins are unpardonable (F).

4. *Moral posture.* 410. I would certainly enjoy beating a crook at his own game (T). 181. When I get bored, I like to stir up some excitement (T). 94. I do many things which I regret afterwards (I regret things more or more often than others seem to) (F). 253. I can be friendly with people who do things which I consider wrong (T). 109. Some people are so bossy that I feel like doing the opposite of what they request, even though I know they are right (T). 208. I like to flirt (T). 430. I am attracted by members of the opposite sex (T). 548. I never attend a sexy show if I can avoid it (F). 231. I like to talk about sex (T). 378. I do not like to see women smoke (F). 355. Sometimes I enjoy hurting persons I love (T).

5. *Sense of reality.* 33. I have had very peculiar and strange experiences (F). 349. I have strange and peculiar thoughts (F). 251. I have had blank spells in which my activities were interrupted and I did not know what was going on around me (F). 48. When I am with people, I am bothered by hearing very queer things (F). 22. At times I have fits of laughing and crying that I cannot control (F). 192. I have had no difficulty in keeping my balance in walking (T). 62. Parts of my body often have feelings like burning, tingling, crawling, or like "going to sleep". 541. My skin seems to be unusually sensitive to touch (F).

6. *Personal adequacy, ability to cope.* My plans have frequently seemed so full of difficulties that I have had to give them up (F). 82. I am easily downed in an argument (F). 32. I find it hard to keep my mind on a task or job (F). 244. My way of doing things is apt to be misunderstood by others (F). 555. I sometimes feel that I am about to go to pieces (F). 544. I feel tired a good deal of the time (F). 261. If I were an artist, I would like to draw flowers (F). 554. If I were an artist, I would like to draw children (F). 132. I like collecting flowers or growing house plants (F). 140. I like to cook (F). 380. When someone says silly or ignorant things about something I know, I try to set him right (T).

7. *Phobias, infantile anxieties.* 367. I am not afraid of fire (T). 525. I am made nervous by certain animals (F). 510. Dirt frightens or disgusts me (F). 494. I am afraid of finding myself in a closet or small closed place (F). 559. I have often been frightened in the middle of the night (F).

8. *Miscellaneous.* 221. I like science (T). 513. I think Lincoln was greater than Washington (T). 561. I very much like horseback riding (F). 458. The man who had most to do with me when I was a child (such as my father, stepfather, etc.) was very strict with me (T). 421. One or more members of my family is very nervous (T).

The pretherapy characteristics of patients who improve in therapy, as compared with those who do not improve, might be summarized as follows:

Improved: (*a*) good physical functioning; (*b*) spontaneity, ability to share emotional experiences; (*c*) conventional church membership, but nonfundamentalist and undogmatic in religious beliefs; (*d*) permissive morality; (*e*) good contact with reality; (*f*) feelings of personal adequacy and vitality; (*g*) physical courage and lack of fear.

Unimproved: (*a*) many and chronic physical ailments; (*b*) broodiness, inhibition, a strong need for emotional seclusion, worrisomeness; (*c*) intense religious experiences, belief in prayer, miracles, the Bible; (*d*) repressive and punitive morality; (*e*) dissociation and ego-alienation; (*f*) confusion, submissiveness, chronic fatigue; (*g*) phobias and infantile anxieties.

From an inspection of these differences, one might easily be led to envy the mental salubrity of psychoneurotic patients who are about to improve. Their actual mental distress, however, has been detailed in case material presented in two previous reports [2, 3], and will not be repeated here. What the group comparison really reveals, of course, is the *dimension* on which the improved and unimproved groups differed. Had the improved patients been compared with an exceptionally healthy nonclinic group of subjects, the same items might well have emerged as descriptive of the difference between the groups, but with the characteristic responses of the improved patients being exactly opposite to those listed above. In other words, the nature of the criterion behavior determines the *nature* of the dimension which the item analysis will reveal, but the question of the *strength* of that variable in the criterion groups must be answered separately.

In this case, it is suggested that what is being measured is a general factor of capacity for personality integration, or ego-strength. The greater vividness of psychopathology often tends to obscure the ego-synthetic or constructive forces in the behavior of a psychologically disturbed individual, so that a prognostic evalution is generally more difficult to make than a diagnostic evaluation. Nevertheless, in spite of the saliency of psychopathology in the clinical picture, it may be presumed that the patient has certain latent strengths which will gradually show themselves, particularly as the psychological crisis which brings him to therapy subsides. What the item content of the prediction scale seems to indicate is that these strengths are of the sort that are generally ascribed to a well-functioning ego, and that it is latent ego-strength which is the most important determinant (within the patient) of response to brief psychotherapy.

Such an interpretation would, of course, have relatively little warrant without supporting evidence from other samples. The obvious next step is to inquire into the relation of the scale to other measures in new populations.

Correlates of the Prediction Scale

In this further inquiry, one clinic sample and two nonclinic samples were studied. The clinic sample consisted of 77 women and 50 men who were seen for diagnostic studies on the Psychology Service at Langley Porter Clinic during a given term of work. The nonclinic samples consisted of 160 male Air Force officers and 40 male graduate students.

The first step was to obtain adjective descriptions, by objective and skilled observers, of high and low scorers on the prediction scale. This was made possible by administering the MMPI to graduate students who were participating in intensive three-day psychological assessments being conducted at the Institute of Personality Assessment and Research, University of California, Berkeley. Following the assessment periods, the staff members of the Institute filled out an adjective checklist with the purpose of describing each one of the subjects in the assessment, on the basis of the subject's socially observable behavior in situational procedures, interviews, and informal social interaction. A composite staff impression was thus assembled. The 10 highest and 10 lowest scorers on the psychotherapy prediction scale were then compared by item-analyzing the composite adjective list for the two groups. The adjectives which showed a statistically significant difference (.05 level) between high and low scorers are listed below. (The staff observers were, of course, in ignorance of the test scores of the subjects.)

a. Adjectives checked more frequently about high-scoring subjects: alert, adventurous, determined, independent, initiative, outspoken, persistent, reliable, resourceful, responsible.

b. Adjectives checked more frequently about low-scoring subjects: affected, dependent, effeminate, mannerly, mild.

The general impression conveyed is of greater resourcefulness, vitality, and self-direction in the high scorers, and effeminacy, inhibition, and affectation in the low scorers. This picture is supported by staff ratings of these same subjects on a number of psychological variables which it was thought could be inferred from social behavior in the assessment setting. The psychotherapy prediction scale correlated significantly with Vitality (.38), which was defined simply as "general energy level," and with Drive (.41), defined as "persistence, resolution, perseverance, *directed* energy." In addition, the scale showed low but positive correlation with several other variables descriptive of effective functioning. These are: Self-confidence (.24), Poise (.24), and Breadth of Interest (.25). Significant negative correlations are with Submissiveness (−.40), Effeminacy (−.34), and Intraceptiveness (−.34).

As in the adjective descriptions, the high scorers on the scale emerge as more adequate physically, more at ease socially, and somewhat broader culturally. Low-scoring men are effeminate, submissive, and inclined to turn inwards rather than to be emotionally outgoing.

The relationship of the prediction scale to intelligence was next investigated. Among the functions of the ego, as described by psychoanalytic writers and summarized by Fenichel [4], are perceiving, planning, synthesizing, and, in general, bringing the subject into an adaptive relationship to reality. Ego-determined behavior is what we are accustomed to calling *intelligent* behavior. Any scale which purports to measure ego-strength should be positively correlated with standardized measures of intelligence.

In the sample on which the scale was developed, its correlation with Wechsler-Bellevue IQ is .44. In the Air Force officer sample, the scale correlates .36 with total score on the Primary Mental Abilities test, and .47 with the Intellectual Efficiency scale of the California Psychological Inventory. In that same sample it correlates .48 with the Potential Success scale developed at the Institute of Personality Assessment and Research against a criterion consisting of faculty ratings of the probable professional success of the doctoral candidates studied. In the graduate student sample itself, the psychotherapy prediction scale correlates .39 with intelligence as measured by the Miller Analogies Test. Further, in the latter sample it correlates .52 with the Intellectual Efficiency scale. Thus it relates to general intelligence, as measured by a variety of tests, even in highly restricted ranges of intelligence. Certainly the ego-strength interpretation of the scale is supported.

The scale also is related to tolerance and lack of ethnic prejudice. In the standardization sample it correlates −.47 with the Ethnocentrism (*E*) scale of Form 60 of the University of California Public Opinion Study Questionnaire. In the graduate student sample it correlates −.35 with the Prejudice scale [5] of the MMPI, and −.46 with the *E* scale. Its correlation with the *E* scale in the Air Force officer sample is −.23, and its correlation with the Tolerance scale of the CPI in that sample is .42.

Again, these findings lend some weight to the notion that what is being measured is general excellence of ego-functioning. The authors of *The Authoritarian Personality* [1], in their successful search for the character defects which accompany ethnocentrism, found that high scorers on the *E* scale show "lack of differentiation of the ego." This is manifested among clinic patients [7], in "a narrow range of experience, emotionally and intellectually," together with "rigidity and constriction," "stereotyped thinking," and so on. All of these things are thus, by inference, negatively related to scores on the psychotherapy prediction scale.

The relationship of the scale to the diagnostic and validity scales of the MMPI was determined in both the graduate student sample and the diagnostic study cases. The results are shown in Table 1. Surprisingly high negative correlations are found with most of the measures of psychopathology, averaging in the neighborhood of −.60 with Hypochondriasis, Depression, Hysteria, Psychasthenia, and Schizophrenia, and around —.50 with Paranoia in the clinical samples. What this suggests is that the prediction scale is picking up a general factor of psychopathology in the MMPI, reflecting degree of maladjustment or ego-dysfunction irrespective of differential diagnosis. In other words, it is related to general elevation of the profile, regardless of the pattern.

Table 1
Relationship of Ego-strength Scale to Diagnostic and Validity Scales of the MMPI in Clinic and Student Populations

Scale	A. Male Clinic Patients ($N=50$)	B. Male Graduate Students ($N=36$)	C. Female Clinic Patients ($N=77$)
F	−.49	−.36	−.47
K	.31	.31	.31
Hs	−.62	−.67	−.63
D	−.60	−.53	−.67
Hy	−.39	−.61	−.63
Pd	−.48	−.07	−.34
Mf	−.04	−.43	.07
Pa	−.62	−.07	−.49
Pt	−.71	−.54	−.71
Sc	−.55	−.44	−.64
Ma	−.04	−.33	−.21

The correlations are, of course, partly a function of overlap of items (generally scored in the reverse direction) between the prediction scale and the diagnostic scales. The amount of net overlap for each scale is shown in Table 2. The very fact of overlap itself testifies to the character of the scale as a measure of general excellence of ego-functioning, manifested in this instance by absence of chronic psychopathology.

Table 2

Item Overlap of Ego-strength Scale and MMPI Diagnostic Scales

Scale	Items Scored in Same Direction	Items Scored in Opposite Direction	Net Overlap
Hs	None	2, 43, 51, 62, 153, 189, 192	Seven items, scored opposite
D	58, 241	2, 32, 36, 43, 51, 153, 189, 208, 236, 270	Eight items, scored opposite
Hy	253	2, 32, 43, 51, 109, 153, 174, 189, 192, 234	Nine items, scored opposite
Pd	82	32, 33, 94, 231, 244	Four items, scored opposite
Pa	None	109, 341	Two items, scored opposite
Pt	None	22, 32, 36, 94, 189, 217, 344, 349, 359	Nine items, scored opposite
Sc	None	22, 187, 192, 241, 251, 349	Six items, scored opposite
Ma	109, 181	22, 100, 251	One item, scored opposite

The scale has not yet been put in a *T*-score form which would be comparable to the profile scores of other MMPI scales. This is principally because no representative sample of normal records has been available to the writer. The raw score means and standard deviations of the samples studied are presented in Table 3, however. The values are fairly consistent in the clinical samples, with somewhat higher means for the graduate students and Air Force officers.

It would be justifiable, of course, to determine *T* scores on the basis of the clinic samples alone, provided that the scale is to be interpreted solely as a Psychotherapy Prognosis scale for use in an outpatient clinic. In Table 4, a chart is given for conversion of raw scores into *T* scores when the scale is used only for such prediction purposes rather than as a measure of ego-strength in the general population.

Table 3

Ego-strength Scale Statistics

Sample	*N*	Range	Mean	SD
1. Standardization sample	33	15–62	41.94	13.30
2. 1st cross-validation sample*	53	26–58	40.10	7.62
3. 2nd cross-validation sample	52	23–60	41.04	8.18
4. 3rd cross-validation sample	46	22–61	42.06	9.32
5. Clinic diagnostic cases	127	23–60	41.97	7.36
6. VA mental hygiene clinic patients	52	22–59	41.79	7.38
7. I.P.A.R. graduate students	40	37–60	50.92	5.62
8. Air Force officers	160	38–60	52.73	4.05

* Scores on abbreviated scale prorated for length of scale.

Table 4

Conversion Table for Use of Ego-strength Scale in Clinic Samples

Raw Score	*T* Score	Raw Score	*T* Score
22	25	46	57
23	27	47	58
24	29	48	59
25	30	49	60
26	32	50	62
27	33	51	63
28	34	52	64
29	36	53	65
30	37	54	67
31	38	55	68
32	39	56	69
33	40	57	70
34	42	58	72
35	43	59	73
36	44	60	74
37	45	61	75
38	47	62	76
39	48	63	78
40	49	64	79
41	50	65	80
42	52	66	82
43	53	67	83
44	54	68	85
45	56		

Cross-validation of the Scale as a Prediction Instrument

The prediction scale was tested on three new samples of psychotherapy patients, in order to see whether it was doing the job for which it was designed. The patients were psychoneurotic, the psychotherapy was brief (very close to six months in all cases), and the clinical setting was similar to that in which the scale was developed. All of the patients took the MMPI at the beginning of therapy, and all were rated as to degree of improvement at some date following termination of therapy. The three cross-validating samples are described below.

1. Fifty-three patients who were given psychotherapy because of delayed recovery from injury or physical disease. Ratings of improvement were made by expert judges (who had not taken part in the therapy) on the basis of terminal interviews. At a final rating conference, these 53 patients were ranked in terms of degree of improvement. This sample of patients had been studied some six years earlier by Harris and Christiansen [6], who showed that patterns of test scores on both the Rorschach and the MMPI were predictive of improvement. The same pretherapy MMPI's were now scored on the prognosis scale. Only 39 items of the 68 could be used, however, as these patients had been given a shortened form of the MMPI, containing only the items which make up the clinical and validity scales of the test. The correlation with terminal rating is .42.

2. Fifty-two patients who had received brief psychotherapy during the preceding five years at Langley Porter Clinic. The sample was obtained by asking therapists who had worked at the clinic during that period of time to nominate patients whom they remembered as clear examples of exceptional improvement, complete lack of improvement, and moderate improvement. The latter group consisted of 27 patients, while the two extreme groups numbered 9 and 16 respectively. The degree of relationship between pretherapy prognosis scale scores and this trichotomy, as determined by computation of *eta*, is .54. The means were as follows: Unimproved, 32.75; Improved, 43.07; Exceptional Improvement, 49.66.

3. Forty-six patients who were part of the current patient load on the psychiatric service of a general hospital. All of these patients had had approximately six months of psychotherapy at the time the rating was obtained. Ratings were made by the therapists themselves, on a 9-point scale of improvement. The correlation with pretherapy prognosis scale scores is .38.

These correlation coefficients are of about the magnitude that one would expect for a scale which is giving a valid measure of patient variables related to outcome of psychotherapy. It is reasonable that there should remain considerable unaccounted-for variance, quite apart from whatever error variance is contributed by the fallible criterion and predictor measure; for there are many important determinants of outcome of therapy besides the personality of the patient. The personality of the therapist, e.g., is also important; and, in addition, there is a subtle interactional factor which results from the combination of a particular patient with a particular therapist, which is not infrequently the crucial determinant of outcome in a given case. Then there are, of course, life-situational variables outside of the therapy, affecting both patient and therapist, individually or jointly. It seems safe to say that no standardized test is likely to achieve any very high order of correlation with therapeutic outcomes. At the same time, the effort to construct measures of the different sources of variance tends to advance the research problem somewhat, and to make new questions answerable. If the scale presented here continues to prove effective upon further investigation, it opens the way to an inquiry into the effect of the therapist's personality, intelligence, and social attitudes upon therapeutic results, with the effect of patient variables held constant. Beyond that question, of course, lies the more difficult problem of measuring interactional effects, and their unique influence on the therapeutic relationship.

Summary and Conclusions

To recapitulate: an MMPI scale was developed for the purpose of predicting response to psychotherapy, and upon inspection of its item content and its personality and intelligence test correlates it was interpreted as being essentially a measure of ego-strength. It is here proposed that the scale be known as the Ego-strength scale of the MMPI, and that the conventional abbreviation for it should be *Es.*

The relationship of the scale scores to therapeutic outcomes in several cross-validating samples led to the conclusion that a significant determinant of personality change in psychotherapy is strength of the ego before therapy begins. Among the characteristics which are collectively referred to as ego-strength are physiological stability and good health, a strong sense of reality, feelings of personal adequacy and vitality, permissive morality, lack of ethnic prejudice, emotional outgoingness and spontaneity, and intelligence.

Since the patients who seek psychotherapy are almost invariably in some sort of psychological difficulty, it must be evident that these characteristics are usually not salient features of the clinical picture at first contact. The evidence suggests, however, that such strengths are latent in the personality, and that they emerge as therapy progresses. By implication, it seems probable that the kind of personal crisis which brings the person of good ego-strength to the clinic is more situation-linked and less characterologically based (i.e., less chronic) than the personal difficulties of the person with poor ego-strength.

The scale should be useful both as a research instrument and as an additional clinical indicator on the MMPI. The present writer would suggest, however, that considerable caution should be exercised in the clinical use of the scale. Certainly it should not serve as the basis for categorical recommendations to treat or not to treat certain patients; the grounds for such action in a clinic should properly involve values as well as facts, and in any event the kind of crude measuring device presented here represents a fairly low order of "fact." Any prognostic assertions made on the basis of this scale should be quite tentative, and probably should be accompanied by a visual image of the kinds of scatter plots which may give rise to correlations in the general neighborhood of .45.

As a research instrument, the scale should prove useful in giving some assessment of the role of "patient variables" in determining the complex outcome which is involved in response to psychotherapy. It may also be of some value in assessing the kind of change that occurs in therapy. One may ask, for example, whether there is actually an enhancement of ego-strength as a consequence of therapy, and get an answer by comparing pretherapy with posttherapy scores on the scale. Or one might inquire whether the therapy itself is causing the change for the better in patients who improve; evidence on the issue might be obtained by setting up a research design which would use as a control group a sample of patients who are matched with the therapy sample in terms both of *Es* scores and of present need for psychotherapy, but who differ in that they do not receive psychotherapy.

In addition, the scale may be useful as an assessment device quite apart from the clinical situation. Its correlates with personality variables in normal samples are similar to the pattern of relationships seen in clinic samples, and in general it seems to be measuring constructive forces in the personality. Thus it may serve as a predictor in any situation in which an estimate of personal adaptability and resourcefulness is called for.

Received December 11, 1952.

References

1. Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D., & Sanford, R. N., *The authoritarian personality.* New York: Harper, 1950.
2. Barron, F. Psychotherapy as a special case of personal interaction: prediction of its course. Unpublished Ph.D. thesis, Univer. of Calif., Berkeley, 1950.
3. Barron, F. Some test correlates of response to psychotherapy. *J. consult. Psychol.,* 1953, **17**, 235–241.
4. Fenichel, O. *Psychoanalytical theory of the neuroses.* New York: Norton, 1945.
5. Gough, H. G. Studies of social intolerance: I. Some psychological and sociological correlates of anti-Semitism. *J. soc. Psychol.,* 1951, **33**, 237–246.
6. Harris, R. E., & Christiansen, C. Prediction of response to brief psychotherapy. *J. Psychol.,* 1946, **21**, 269–284.
7. Levinson, M. H. Psychological ill-health in relation to potential fascism: a study of psychiatric clinic patients. Chapter XXII in [1].

Journal of Consulting Psychology
Vol. 17, No. 5, 1953

MMPI Changes Concomitant with Client-Centered Therapy[1]

James J. Gallagher
Michigan State College

The problem of how to evaluate psychotherapy seems often to be complicated by the larger problem as to what the goals in therapy are, or should be. Nowhere is this problem more evident than in the recent article by Eysenck [3] who reviewed the literature on evaluation of psychotherapy and came to the tentative conclusion that the available evidence fails to support the notion that psychotherapy facilitates recovery from neurotic disorders. To reach this interesting conclusion, Eysenck had to accept as equal or nearly equal the judgments of a great number of psychiatrists and physicians as to what constitutes *cure* or *improvement*. He is probably right in stating that his conclusions will be looked on with a skeptical eye, although he assumes that it will be the result of an emotional block on the part of the skeptics rather than any methodological weakness leading to his own conclusions. It should be clear by now that "success in psychotherapy" is a value judgment concept, a fact which makes it difficult to find a definition acceptable to everyone. The real problem to the science of psychology is that the many discussions and vituperative arguments as to what therapy should be have not advanced us too much further in our search of what therapy is.

The writer would suggest that there is another and more profitable way of approaching this problem, namely, to demonstrate empirically what changes do take place as a result of a particular type of therapy. Thus, objective evidence will be available, regardless of whether or not it is acceptable to various critics, as to what a therapy does do as distinguished from what its followers feel it should do.

Problem

The present problem is to determine whether there were any changes on the results of the MMPI after a group of college students had undergone client-centered therapy. All of the cases, predominantly college students, who came to the Psychological Clinic for aid in their personal adjustment were minutely examined by the group research project in psychotherapy. Interviews were recorded and transcribed, pretests and posttests were administered, therapists' ratings and various judges' ratings were obtained, etc. The present study deals with the MMPI results and their relationship with other measures of progress in treatment.

Previous studies. Since the MMPI has been a popular diagnostic instrument, research has been mainly confined to differential diagnostic studies [6, 8, 9, 10, 11, 15], although some investigations have been completed on MMPI changes as a result of various shock therapies [1, 5, 7, 13].

It has, as yet, seen limited use as a tool for evaluation of psychotherapy. Rashkis and Shaskin [14] used the MMPI to evaluate group psychotherapy of a permissive nature. Using a T score of over 70 as a measure of deviation from the normal on 22 psychiatric battle casualities and 15 anxiety state patients, they found the greatest changes in the direction of improvement in the D, Hy, Hs, and Pt scales. The Pd scale remained unchanged and the Ma scale showed a tendency to increase.

Schofield [16], reasoning that a measure of personality deviation should also measure changes in therapy, compared four groups of women undergoing different types of therapy with pretherapy and posttherapy MMPI scores. He also attempted to develop a scale of items from the MMPI which would predict amenability to therapy.

---

[1] The data for the present sudy were obtained as part of a group research project in psychotherapy at the Pennsylvania State College. The coordinated portion of the research is being published under the title: *Group report of a program of research in psychotherapy, Psychotherapy Research Group.* Pennsylvania State College, in press.

There were two groups of neurotic women, one treated on an outpatient basis and the other hospitalized; and two groups of psychotic women, one group having electroshock treatment and the other having no shock treatment. These groups were different from the population used in the present study in that they had limited education, average intelligence, were older, and all were women.

The outpatient neurotic group that was being treated by junior medical students showed no significant changes on posttherapy MMPI's. The hospitalized psychoneurotics showed significant differences on the *Hs, Hy,* and *Pd* scales and a general lowering effect on the whole profile. The hospitalized psychotics who had no shock treatment showed no significant changes on the posttherapy tests, while the psychotics who underwent shock treatment showed changes on the *F, Hs, D, Hy, Pd, Pt,* and *Sc* scales. A group of nonhospitalized normal women showed no significant changes on tests which were given with the same general time lapse as those of the therapy groups.

Schofield's attempt to establish a predictive scale for therapy was not successful, as there was too much overlap on the scale between the much improved and the slightly improved group. He attributed his lack of success to a "lack of a reliable clinical criterion of patients' responses to therapy."

Wiener and Phillips [18] studied the changes of the MMPI on a patient undergoing modified psychoanalysis (one hour a week). During a period of eight months there was noticed a strong drop in the *D, Pt,* and *Sc* scales and a moderate drop in the *Hy* and *Pa* scores. It was also noted that the *K* score went up and that these changes remained when the patient was re-examined a year after therapy was discontinued.

Mosak [12] reported the results of 28 clients who had undergone client-centered psychotherapy for an average number of fifteen interviews. He reported that the posttherapy MMPI results indicated significant decreases on the *D, Sc, Hs, Hy,* and *Pa* scales. The individual profiles showed the same pattern as was observable on the pretherapy tests with a general drop in scores on all the scales. The Bell Adjustment Inventory, also given to the same clients, showed significant improvement in the feelings toward the self and toward others.

Hypotheses. The following hypotheses were tested in the present study.

1. Changes that take place as a result of any type of psychotherapy should be reflected in subsequent changes in the results of personality tests. As there is a change in the direction of health in the individual, there should be a corresponding depression in the scores that represent maladjustment.

2. It was expected that the greater the move to health in the individual, the greater the expected change in the personality tests toward health. Thus, a positive correlation between change in maladjustment scores on the MMPI and the criterion measures of success in psychotherapy was expected.

Method

Subjects

The 41 subjects involved in this study were all students of the Pennsylvania State College who came to the Psychological Clinic between September 1949 and August 1950, either by request of various agencies or upon their own volition, to obtain aid in their personal adjustment. All these subjects were counseled by advanced graduate students in clinical psychology who had been trained in client-centered therapy. No attempt was made to diagnose these cases prior to therapy except that persons judged as prepsychotic or psychotic on the basis of initial contacts were transferred to the staff psychiatrist. The median number of therapy interviews was between five and six, although some cases were seen for a much longer period of time.

Figures obtained from the registrar of the college for the year 1949-1950 showed that the group who came for therapy was representative of the total college population for the semester enrolled, in age (mean = 21.75) and sex (75% males). There was a higher proportion of liberal arts students in the therapy group than in the general college population. No evidence was available to suggest why this was so.

Raw MMPI scores of 202 randomly selected college students were used as a typical college normative group with which to compare the therapy population. The mean raw scores for each of the MMPI scales was obtained and then converted into a *T* score so that some comparison would be possible. It would have been desirable to have a control group of maladjusted students who did not undergo psychotherapy. However, practical considerations militated against such a procedure at this time.

Procedure. A battery of tests consisting of the Rorschach, MMPI, and the Mooney Problem Check List was given to each of the subjects before the beginning of the therapy interviews or, at the latest, before the second therapy interview. The posttherapy tests were given after an agreement was reached by the therapist and client that the therapy contacts should end.

In order to correlate the change on the MMPI with other measures of success, a maladjustment scale was developed to give one score which would represent the amount of deviance from the normal in each case. To obtain this score the deviant scores on all the items on the *Hs, D, Hy, Pd, Pa, Pt,* and *Sc* scales were summed. The *Mf* and *Ma* scales were not included because norms on college students revealed these scales to be unduly elevated. The procedure of summing the individual items was done in preference to summing the *T* scores, since some items appear on more than one scale and would be unduly weighted by such a procedure.

The therapy criterion measures. A multiple criterion for evaluating client-centered psychotherapy with college students was developed by Tucker [17] using the cases in the present group. This multiple criterion consisted of four measures:

1. A 29-item Therapy Rating Scale scored by the therapist following the end of contacts with the client.
2. The same Therapy Rating Scale scored by judges who read at least 60% of the transcribed interviews including the first and last interviews.
3. A 14-item Client Rating Scale which the client filled out at the time he took the posttherapy test battery.
4. The ratio of negative to positive feelings in the final interview as compared to the ratio of negative to positive feelings in the first interview.

These four criterion measures were combined into a total criterion score by weighting each measure according to the ratio of its summed intercorrelation with the other three measures. All of the rating scales were constructed especially for this study by members of the group research project. For further information concerning the reliability of judges' ratings, the construction of rating scales, etc. the reader is referred to Tucker's original work [17].

Table 1

Pre- and Posttherapy MMPI Scores on Therapy Group Compared with Random Sample Group of College Students

MMPI Scale		Pretherapy Mean *T* score $N = 41$	Posttherapy Mean *T* score $N = 41$	Random Sample Mean *T* score $N = 202$
F	Mean	57.92	55.45**	57†
	Sigma	7.73	6.91	9
K	Mean	53.05	55.34**	53
	Sigma	6.45	8.89	8
Hs	Mean	55.17	51.71**	47
	Sigma	12.85	9.28	4
D	Mean	67.10	60.88*	50
	Sigma	12.51	12.54	6
Hy	Mean	60.66	59.54	56
	Sigma	10.89	8.56	6
Pd	Mean	62.49	61.17	53
	Sigma	13.48	10.03	9
Pa	Mean	56.67	55.22	53
	Sigma	8.29	9.33	8
Pt	Mean	67.49	63.49*	47
	Sigma	8.52	11.83	14
Sc	Mean	64.43	62.20	49
	Sigma	12.72	11.68	11
Ma	Mean	57.17	58.97	57
	Sigma	7.98	7.70	11
Si	Mean	59.94	55.03*	----
	Sigma	11.32	10.02	----
Mal.	Mean‡	86.56	79.41*	----
	Sigma	17.51	18.04	----

‡ Mean raw scores.
† Estimated T scores on the basis of raw score data.
* Difference between pretherapy and posttherapy significant at 1% level of confidence.
** Difference between pretherapy and posttherapy significant at 5% level of confidence.

Results

1. Table 1 shows significant differences between pretherapy and posttherapy mean *T* scores on the *F, K, Hs, D, Pt,* and *Si* scales. There was also a significant difference between groups on the maladjustment scale constructed especially for this study. These mean changes include, of course, a great variety of individual changes. Some cases showed a 30 to 40 *T*-score improvement on certain scales while others, surprisingly enough, showed an upward surge in their profiles. It was the opinion of most therapists handling such cases that such an upward surge was indicative of a breakdown of defenses and the individual's greater awareness of his problems, rather than an indication of a sudden turn to greater maladjustment.

2. All the scales, with the exception of *K* and *Ma,* showed a tendency to be lower on the posttherapy test. This finding agrees with Mosak, who reported a general lowering of all scales as a result of therapy.

3. Despite a considerably lower mean score on posttherapy results, the mean *T* scores of the therapy group were still considerably higher than the random sample group of college students. This would seem to suggest that a sizable amount of improvement might still be profitably made with the therapy group. It must be remembered that the posttherapy tests were given immediately after therapy had been completed. There seems to be some evidence [2] to suggest that further change and improvement in the client takes place for a period after therapy has been completed.

4. The scales generally regarded as the character or behavior disorder scales, *Hy, Pd,* and *Ma,* showed the least tendency to change as a concomitant of client-centered therapy. Conversely the feeling or discomfort scales, *D, Pt,* and *Hs,* showed the greatest tendency towards change. It should be noted that the character and behavior disorders have been traditionally the group considered least susceptible to any kind of psychotherapy and this may be more evidence supporting that clinical finding, rather than a specific failing of client-centered therapy.

Table 2

Correlation of MMPI Maladjustment Change Score with Client-Centered Therapy Criteria †

Criterion measure	Maladjustment change score
Therapist Rating	.15
Judge Rating	.33
Client Rating	.58*
P-N Feeling Ratio	.44**
Multiple Criterion	.41**

† N = 30.
* Significant at 1% level.
** Significant at 5% level.

5. Table 2 shows the correlation of MMPI maladjustment change scores with the various criterion measures developed by Tucker. Although there is a positive correlation with all four criterion measures, only the Client Rating Scale and the Positive-Negative Feeling Ratio showed significant correlations with the Maladjustment Change Scores. This would seem to represent another example of the phenomenon mentioned previously by Tucker [17] and Gallagher [4], that the client's evaluation of his progress and the psychologist's judgment of the client's progress are fairly independent of each other. Since the MMPI is, by and large, a measure of self-evaluation, we would expect fairly high correlations between MMPI changes and Client Rating Scales and Positive-Negative Feeling Ratios. Similarly, we might expect low correlations between the therapist and judge's ratings and the MMPI change scores, and this is what happened.

Discussion

The results of the present study seemed to indicate that there were some changes, as measured by the MMPI, taking place in a group of college students concomitant with client-centered therapy. It should be noted also that considerable changes took place despite two unfavorable factors present in this study: first, the use of relatively inexperienced therapists, and second, a shorter number of interviews than is generally accepted as necessary to accomplish real change in a client.

Also of interest was the fact that some scales were more susceptible to change than others. Generally it was the mood or feeling scales which showed the greatest change in the direction of health, while the behavior or character disorder scales showed the least amount of change toward health. While this is generally true of all types of psychotherapy, consistent results along this line would suggest that those patients who show initially high scores in the behavior disorder scales might profitably be handled in some other fashion than by being submitted to psychotherapy, in this case, client-centered psychotherapy.

Another finding of interest was that there seem to be two different and relatively independent frames of reference towards success in psychotherapy; the client's frame of reference in which he estimates change that has occurred within himself, and the therapists' frame of reference in which they estimate the amount of change that has taken place within the individual. While we, as psychologists, would like to believe that our expert opinions are more accurate and more realistic in the evaluation of actual change, it does not necessarily follow that this is so. Some further study in which the opinion of both client and therapist are checked against some observable behavioral change in the client would do much to clear up this problem.

The existence of such different frames of reference would pose a particular problem for client-centered therapy, for if the client is to have the final and irrevocable decision as to when he will leave therapy, he may often leave a rather frustrated therapist behind, who feels that the client has escaped therapy with only some anxiety reduction rather than a permanent solution or adaptation to his problem. Certainly some effort needs to be made to bring these frames of reference into closer conjunction with each other.

Summary

Forty-two college students who underwent client-centered therapy from advanced graduate students in clinical psychology were given a pretherapy and posttherapy battery of tests, one of which was the MMPI. The purpose of the present study was to determine if there were any changes in the MMPI concomitant with the application of therapy and to see how well the changes on the MMPI compared with a multiple criterion of success in therapy.

The results showed significant differences in the direction of health on six of the MMPI scales with the greatest differences being shown on the feeling or mood scales, while the least differences were shown on the character and behavior disorder scales. The MMPI maladjustment change scores showed significant correlations with measures of self-rating of success in therapy and of change in ratio of positive and negative feelings from beginning to end of therapy.

Received January 23, 1953.

References

1. Carp, A. L. MMPI performance and insulin shock therapy. *J. abnorm. soc. Psychol.*, 1950, **45**, 721–726.
2. Cowen, E. L., & Combs, A. W. Follow-up study of 32 cases treated by nondirective therapy. *J. abnorm. soc. Psychol.*, 1950, **45**, 232–258.
3. Eysenck, H. J. The effects of psychotherapy: an evaluation. *J. consult. Psychol.*, 1952, **16**, 319–324.
4. Gallagher, J. J. A comparison of clients who continue with clients who discontinue client-centered therapy. In *Group report of a program of research in psychotherapy, Psychotherapy Research Group*. Pennsylvania State College, in press.
5. Hales, W. M., & Simon, W. MMPI patterns before and after insulin shock therapy. *Amer. J. Psychiat.*, 1948, **105**, 254–258.
6. Hampton, P. The MMPI as a psychometric tool for diagnosing personality disorders among college students. *J. soc. Psychol.*, 1947, **26**, 99–108.
7. Harris, R. E., Bowman, K. M., & Simon, A. Studies in electronarcosis therapy. III. Psychological test findings. *J. nerv. ment. Dis.*, 1948, **107**, 371–376.
8. Hunt, H. F., Carp, A. L., *et. al.* A study of the differential diagnosis efficiency of the MMPI. *J. consult. Psychol.*, 1948, **12**, 331–336.
9. Lough, Orpha. Teachers college students and the MMPI. *J. appl. Psychol.*, 1946, **36**, 241–247.
10. Meehl, P. E. Profile analysis of the MMPI in differential diagnosis. *J. appl. Psychol.*, 1946, **30**, 517–524.
11. Morris, W. W. A preliminary evaluation of the diagnostic scales of the MMPI. *Amer. Psychologist*, 1946, **1**, 264. (Abstract)
12. Mosak, H. H. Evaluation of psychotherapy: a study of some current measures. Unpublished doctor's dissertation, Univer. of Chicago, 1950.
13. Pacella, B. L., Piotrowski, Z. A., & Lewis, N. D. The effects of electric convulsive therapy on certain personality traits in psychiatric patients. *Amer. J. Psychiat.*, 1947, **104**, 83–91.
14. Rashkis, H. A., & Shaskin, D. A. The effects of group therapy on personality scores. *Amer. J. Orthopsychiat.*, 1946, **16**, 343–349.
15. Rubin, H. The Minnesota Multiphasic Personality Inventory as a diagnostic aid in a veterans hospital. *J. consult. Psychol.*, 1948, **12**, 251–254.
16. Schofield, W. Changes in responses to the Minnesota Multiphasic Inventory following certain therapies. *Psychol. Monogr.*, 1950, **64**, No. 5 (Whole No. 311).
17. Tucker, J. E. A multiple criterion of success in client-centered therapy. In *Group report of a program of research in psychotherapy, Psychotherapy Research Group*. Pennsylvania State College, in press.
18. Wiener, D. N., & Phillips, E. L. A study of progress in psychotherapy. *J. clin. Psychol.*, 1948, **4**, 201–206.

Journal of Consulting Psychology
Vol. 17, No. 5, 1953

An Adjustment Score for Q Sorts

Rosalind F. Dymond
Counseling Center, University of Chicago

One of the most serious of the difficulties faced by those who would undertake research on the results of psychotherapy is the selection of suitable criteria. The person who presents himself for therapy presumably feels in need of change. Therefore, to measure the effects of the therapy, there must be some method of assessing his status before and after therapy, and some explicit standard against which the change, if any, may be measured. Such a criterion measure would also make possible group comparisons such as the initial adjustment status of an experimental or therapy group in comparison to a no-therapy group, and the comparative standing of these two groups after the experimentals have completed their therapeutic interviews.

To provide one such measure for the program of research in therapy under way at the Counseling Center of the University of Chicago,[1] the *Q*-sort adjustment score was devised. The *Q*-sort instrument itself has already been briefly described in this series [1]. It is a test made up of 100 self-referent statements which each subject was asked to distribute into nine piles, putting a prescribed number of cards into each, making a forced normal distribution. The instructions are that he put the cards most descriptive of him as he sees himself at one end, and the least descriptive at the other, with the items about which he is indifferent or undecided falling at the center. The required distribution is as follows:

	Least like me					Most like me			
Pile no.	0	1	2	3	4	5	6	7	8
No. of cards	1	4	11	21	26	21	11	4	1

The statements to be distributed were compiled by John Butler and Gerard Haigh[2] from therapeutic interviews so as to represent both the positive and negative statements that clients frequently make about themselves. As has been outlined by Grummon [1], both the experimental and control subjects in the research on Block I sorted these cards at various times. To review the design briefly, Block I of this program involved an experimental group of twenty-five cases selected randomly from those presenting themselves for therapy and a control group of equal size of "normals." The control group is normal in the sense that they had never had any psychotherapy and were not at the time of selection so motivated. They were individually matched with the experimental subjects on age, sex, socioeconomic status, and occupation. To control for the possible effect of the passage of time, the tests for the controls were spaced at the same intervals as those of the therapy cases with whom they were matched. This also gives control over the possible effect of repeated testing. There was no attempt to match on such factors as psychological status, personality variables, or motivation for therapy. These were handled by using an "own-control" design as well. The experimental (therapy), and the control (no-therapy) groups were each subdivided into an *A* and a *B* category. The experimental *A*'s were given a battery of tests two months before their therapy began, and again just before their therapy started. This made available a measure of their pretherapy psychological status, and also permitted a comparison of their prewait and pretherapy tests to reveal the extent to which a group motivated to seek therapy would change in a positive direction, in a two-month interval, without the help of therapy. A positive shift between these two sets of test scores in the experimental group during the no-therapy waiting period would indicate that the important variable in "improvement" is not the therapy but the motivation. However, if the control group which is not motivated for therapy were to show a shift of the same magnitude, the change might better be attributed to the practice effect resulting from retesting.

---

[1] This is one of a series of articles reporting a large-scale research program in psychotherapy and personality, which is being conducted by the staff of the Counseling Center of the University of Chicago. This research program is largely supported by grants from the Rockefeller Foundation through its Medical Division. Numbers I, II, and III in the series appeared in the *Psychological Service Center Journal*, 1951, 3, 3–165.

[2] Haigh, G., & Butler, J. Proposal for research on changes in perception of self and others in therapy. In First interim report of the Rockefeller research project. Univer. of Chicago, unpublished manuscript, pp. 12–15.

The experimental *B* group had no waiting period but was first tested just before its therapy began. As the control sample was tested at the same time intervals as its experimental match, for half the sample of both experimentals and controls there were four sets of tests: prewait, pretherapy, posttherapy, and follow-up (six months after the posttherapy tests). For the other half, there were three testing points: pretherapy, posttherapy, and follow-up.

To obtain an adjustment criterion against which any particular person's distribution of the cards could be scored, the entire group of 100 statements was given to two well-trained, practicing clinical psychologists from outside the client-centered orientation. They were asked to sort the statements into two piles, those the well-adjusted person should say are like him, and those the well-adjusted person should say are unlike him. Both judges felt that some items were irrelevant to one's adjustment status, and were permitted to omit these items from their sortings. The two distributions made by the two judges agreed remarkably well, differing on only 2 of the 100 items. Next, the 26 statements which the initial two judges agreed did not differentiate adjustment were discarded, and four new judges were given the remaining 74 to sort. The new judges were asked to make two equal piles, so that there would be 37 items that the well-adjusted individual would say are like him, and 37 he would say are unlike him. Again the agreement among the judges' sortings was very high. Only one judge differed on the placement of as many as four items. The composite picture of the self-description of the well-adjusted person was then tabulated as 37 positive indicators on the "like me" side of the distribution (from scale position 5 to 8), and 37 negative indicators on the "unlike me" side (between 3 and 0). Table 1 lists these items.

Table 1

Adjustment Score Items

Negatives: Contribute to score if fall on "unlike me" side (0–3)

1. I put on a false front.
2. I often feel humiliated.
3. I doubt my sexual powers.
4. I have a feeling of hopelessness.
5. It is difficult to control my aggression.
6. I want to give up trying to cope with the world.
7. I tend to be on my guard with people who are somewhat more friendly than I had expected.
8. I usually feel driven.
9. I feel helpless.
10. My decisions are not my own.
11. I am a hostile person.
12. I am disorganized.
13. I feel apathetic.
14. I don't trust my emotions.
15. It is pretty tough to be me.
16. I have the feeling that I am just not facing things.
17. I try not to think about my problems.
18. I am shy.
19. I am no one. Nothing seems to be me.
20. I despise myself.
21. I shrink from facing a crisis or difficulty.
22. I just don't respect myself.
23. I am afraid of a full-fledged disagreement with a person.
24. I can't seem to make up my mind one way or another.
25. I am confused.
26. I am a failure.
27. I am afraid of sex.
28. I have a horror of failing in anything I want to accomplish.
29. I really am disturbed.
30. All you have to do is just insist with me and I give in.
31. I feel insecure within myself.
32. I have to protect myself with excuses, with rationalizing.
33. I feel hopeless.
34. I am unreliable.
35. I am worthless.
36. I dislike my own sexuality.
37. I have few values and standards of my own.

Positives: Contribute to score if fall on "like me" side (5–8)

1. I am likeable.
2. My personality is attractive to the opposite sex.
3. I am relaxed and nothing really bothers me.
4. I am a hard worker.
5. I feel emotionally mature.
6. I am intelligent.
7. I am self-reliant.
8. I am different from others.
9. I understand myself.
10. I am a good mixer.
11. I feel adequate.
12. I am satisfied with myself.
13. I am assertive.
14. I take a positive attitude toward myself.
15. I have initiative.
16. I am ambitious.
17. I have an attractive personality.
18. I am tolerant.
19. I am a rational person.
20. I am impulsive.
21. I am poised.
22. I am contented.
23. I can usually make up my mind and stick to it.
24. I am sexually attractive.
25. I am liked by most people who know me.
26. I am optimistic.
27. My hardest battles are with myself.
28. I can usually live comfortably with the people around me.
29. I express my emotions freely.
30. I usually like people.
31. Self-control is no problem to me.
32. I can accept most social values and standards.
33. I am a responsible person.
34. I have a warm emotional relationship with others.
35. I often kick myself for the things I do.
36. I am responsible for my troubles.
37. I make strong demands on myself.

Any person's resemblance to this ideal type can be computed by counting the number of the 74 items which he places on the appropriate side of the distribution when describing himself. This tally was called the adjustment score.

Table 2

Adjustment Score of "*A*" Group in the No-Therapy Period

Test	Experimental group ($N=11$) Mean	Experimental group ($N=11$) Range	Control group ($N=11$) Mean	Control group ($N=11$) Range
Prewait	28.7	16–36	47.0	25–58
Pretherapy	31.2	14–41	47.4	27–58

Results

The mean adjustment score of the members of the experimental *A* group at the prewait testing time showed that they described themselves as significantly poorer in adjustment than the no-therapy group, as indicated by the *t* test. The mean adjustment score of neither the experimental nor the control *A* group changed significantly over the waiting period, and the two remained significantly different at the 1 per cent level (Table 2). Thus neither motivation to change nor familiarity with the test seemed to be sufficient in itself to produce a significant increase in adjustment score during the two-month waiting period.

Although the control testing program is not yet complete, a comparison of the 15 completed cases with the 25 experimentals can be made (Table 3).

Table 3

Adjustment Score of Total Group over the Therapy Period

Test	Experimental group ($N=25$) Mean	Experimental group ($N=25$) Range	Control group ($N=15$) Mean	Control group ($N=15$) Range
Pretherapy	28.9	11–42	47.4	27–60
Posttherapy	39.9	17–54	45.6	27–56

In comparing the pre- and posttherapy mean scores, the experimentals are found to increase their adjustment scores significantly over the therapy period, whereas the no-therapy group does not. At the posttherapy testing time, the experimental group mean is not significantly different from the control group mean.

In sum, the group which presumably was motivated to change in some way described themselves as poorer in adjustment than a group that was not so motivated. With the intervention of therapy, this group described themselves as psychologically better in adjustment than previously, and approached the level of a "normal" group without the therapy experience.

Discussion

How good a criterion is the *Q*-sort measure? As this is a self-descriptive measure of adjustment, how well does it agree with a more "objective" measure? Table 4 shows the relation of the Q-sort adjustment score to the counselor's rating of the success of the therapy. Using Yates's exact method for determining the significance of the relationship, it was found to be significant at the 5 per cent level.

Table 4

Relation of Q Adjustment Score of Experimental Group after Therapy and Counselor's Rating of Success

Q Adjustment score	Success rating Low 1–5	 High 6–9
Low (17–38)	7	4
High (39–54)	3	11

It may be argued that these two measures are not independent of each other and that neither is a valid measure of adjustment. The counselor's judgment of the success of the therapy may be influenced by the client's use of well-adjusted self-references in the interviews, and the therapist's attitude of satisfaction with the progress may be communicated to the client who then reflects it in his self-description. There is some possibility that the adjustment score of the follow-up sort may contribute an answer. If the client's posttherapy self-description was greatly influenced by a "need to please" the therapist, or even to prove to himself that he had profited from the time spent in therapy, some of this motivation should have been reduced in the intervening six months. If the gains from therapy were really superficial or minimal and the adjustment score increase wholly spurious, the self-description done six months after therapy might be expected to be somewhat less influenced by the "hello-goodbye effect" [2, p. 228]. The fact that the mean follow-up adjustment score of the therapy group, based on 22 completed cases, is almost identical (38.3) to the posttherapy mean (39.9) seems to indicate some lasting improvement. The follow-up tests of the control group are not yet complete enough to report.

Summary

An index of adjustment based on a Q-sort technique shows a group of persons presenting themselves for therapy to be less well-adjusted, as measured against a criterion set up by expert clinicians, than a group who do not wish therapy. After therapy there is a significant improvement in the experimental group, which is then not significantly different from the no-therapy group. There is, further, a significant concurrence between counselors' judgments of the success of the therapy and the final adjustment status as measured from the self-description.

Received January 26, 1953.

References

1. Grummon, D. L. Studies in client-centered psychotherapy: II: Design, procedures, and subjects for Block I. *Psychol. Serv. Center J.*, 1951, 3, 29–46.
2. Hathaway, S. R. Some considerations relative to nondirective counseling as therapy. *J. clin. Psychol.*, 1948, 4, 226–231.

Journal of Consulting Psychology
Vol. 17, No. 5, 1953

The Appraisal of Body-Cathexis: Body-Cathexis and the Self

Paul F. Secord and Sidney M. Jourard
Emory University

A substantial amount of current research in personality is devoted to the study of the individual's personal world. One object which is ever-present in this personal world is the body. It is the thesis of the present writers that the individual's attitudes towards his body are of crucial importance to any comprehensive theory of personality; yet little attention has been given to this subject by psychologists. The present paper is concerned with one variety of attitude, namely, *body-cathexis.* By *body-cathexis* is meant the degree of feeling of satisfaction or dissatisfaction with the various parts or processes of the body.

If the variable body-cathexis is to be deemed important for personality theory, it is necessary to demonstrate that it is related to other personality variables which are recognized as significant. For reasons which need not be discussed here, body-cathexis is believed to be integrally related to the self-concept, although identifiable as a separate aspect thereof. From this notion of relatedness of body and self, the following more specific hypotheses were formulated and tested:

1. Feelings about the body are commensurate with feelings about the self, when both are appraised by similar scales.
2. Negative feelings about the body are associated with anxiety, in the form of undue autistic concern with pain, disease, or bodily injury.
3. Negative feelings about the body are associated with feelings of insecurity involving the self.

One of the few empirical studies relevant to the present focus is that of Schilder [4], who used a set of questionnaires which probed the feelings, associations, and memories of his patients toward various aspects of their bodies. These semiobjective questionnaires suggested the more objective scale employed in the present study. Another relevant paper, that of Secord [6], describes the use of a homonym word-association technique for the appraisal of body-cathexis in a disguised "projective" fashion. This technique was used here for purposes of comparison with the more direct *BC-SC Scale.*

The BC-SC Scale

The writers approached the problem of appraising body-cathexis (BC) by asking the person to indicate on a scale the strength and direction of feeling which he has about each of the various parts or functions of his body. Such a direct approach, of course, has certain of the theoretical disadvantages which the more typical self-inventories also possess, but it was believed that the technique might have some usefulness in spite of these limitations. The first part of this scale consisted of a listing of 46 body parts and functions. Each item was followed by the numbers 1 through 5. The following instructions appeared on the cover page of the scale:

> On the following pages are listed a number of things characteristic of yourself or related to you. You are asked to indicate which things you are satisfied with exactly as they are, which things you worry about and would like to change if it were possible, and which things you have no feelings about one way or the other.
>
> Consider each item listed below and encircle the number which best represents your feelings according to the following scale:
>
> 1. Have strong feelings and wish change could somehow be made.
> 2. Don't like, but can put up with.
> 3. Have no particular feelings one way or the other.
> 4. Am satisfied.
> 5. Consider myself fortunate.

The second part of the scale concerned self-cathexis (SC), and listed 55 items believed to represent a sampling of the various conceptual aspects of the self, which were rated by the respondents according to the same instructions used for the body scale. The self-traits included were phrased in nontechnical, popular terms, such as *morals, conscience,* and *personality,* so that they might approximate the terms in which the individual actually thinks of himself. Tables 1 and 2 list the items included in both parts.

Table 1

Body-Cathexis Items Used in BC Scale

- hair
- facial complexion
- appetite
- hands
- distribution of hair over body
- nose
- fingers
- elimination
- wrists
- breathing
- waist
- energy level
- back
- ears
- chin
- exercise
- ankles
- neck
- shape of head
- body build
- profile
- height
- age
- width of shoulders
- arms
- chest
- eyes
- digestion
- hips
- skin texture
- lips
- legs
- teeth
- forehead
- feet
- sleep
- voice
- health
- sex activities
- knees
- posture
- face
- weight
- sex (male or female)
- back view of head
- trunk

Table 2

Self-Cathexis Items Used in SC Scale

- first name
- morals
- ability to express self
- taste in clothes
- sense of duty
- sophistication
- self-understanding
- life goals
- artistic talents
- tolerance
- moods
- general knowledge
- imagination
- popularity
- self-confidence
- sensitivity to opinions of others
- ability to lead
- last name
- impulses
- manners
- handwriting
- intelligence level
- athletic skills
- happiness
- creativeness
- love life
- strength of conviction
- conscience
- skill with hands
- ability to express sympathy
- emotional control
- self-consciousness
- generosity
- ability to accept criticism
- thoughts
- artistic & literary taste
- memory
- thriftiness
- personality
- self-respect
- ability to concentrate
- ability to take orders
- fears
- capacity for work
- conscientiousness
- ability to meet people
- self-discipline
- suggestibility
- neatness
- vocabulary
- procrastination
- will power
- self-assertiveness
- ability to make decisions
- dreams

The form of the scale presented in Tables 1 and 2 is the result of considerable preliminary work in which previous forms were tried out on college students. Items which were difficult to understand, difficult for the subject to assign a meaningful rating, or which resulted in little variability from subject to subject were generally eliminated, provided that they did not leave an important part of the body or self unrepresented. One exception to the latter qualification was allowed: organs pertaining to sexual and excretory functions were deliberately omitted from the body list because it was feared that their presence in the scale might give rise to an evasive attitude which would transfer to other items, resulting in an avoidance of the two answer categories representing negative feelings towards the body.

The homonym test of body-cathexis (H test). The H test consisted of a list of 75 homonyms, each of which has meanings pertaining to the body and meanings not related to the body. A substantial proportion of the words have meanings pertaining to pain, disease, or bodily injury. Twenty-five neutral or nonbody words were interspersed with the homonyms for purposes of disguise. This list was presented to the testees orally, the homonyms being read at the rate of one every five seconds, with instructions to the subjects to respond by writing down the first word that occurred to them. A score for each individual was obtained by totaling responses to bodily meanings. It was employed here to provide an independent measure of anxiety-related BC. This test has been fully described elsewhere, and some tentative evidence that it is related to anxiety has been offered [5, 6].

Procedure. The H test and the BC-SC Scale were administered in a group situation to 70 college males and 56 college females. The BC-SC Scale and the Maslow Test of Psychological Security-Insecurity [3] were subsequently administered to an additional group of 47 college men and women.

A study of the patterns of responses of individuals to the BC-SC Scale suggested that response sets were operating for some individuals [1, 2]. In order that split-half reliability coefficients for the BC and SC scores, as well as intercorrelations between these scores, might not be inflated by the artifact of response sets, subjects falling in any of the following arbitrary classes were not included in computations: (*a*) a frequency $\geqq$ 32 in category 4; (*b*) a frequency $\geqq$ 28 in category 5; and (*c*) a frequency $\geqq$ 24 in category 5, when accompanied by less than 2 responses in categories 1 and 2 combined. As a result, the 70 male subjects were reduced to 45, and the 56 females to 43. It should be emphasized that this selection of subjects *lowers* the correlation coefficients obtained between parts of the BC-SC Scale, and *reduces* the split-half reliabilities of the various subscores.

Results

Statistics pertaining to the BC-SC Scale and the homonym test. A single score for each individual was obtained on the H test, consisting of the total number of body responses to the 75 homonyms. Three types of scores were obtained on the BC-SC Scale. These were:

1. Total BC, obtained by summing the ratings for each individual on the 46 body items and dividing by 46.
2. Total SC, obtained by summing the ratings for each individual on the 55 self items and dividing by 55.
3. An *anxiety-indicator* score, obtained by summing the ratings for each male individual on the 11 BC-Scale items most negatively cathected by the group of males: facial complexion, nose, energy level, body build, profile, height, chest, teeth, sex activities, posture, and weight. These sums were divided by 11. A similar score, based on the 11 items most negatively cathected by females, was also obtained.

Table 3

Means, Standard Deviations, and Reliabilities of the BC-SC Scale and the Homonym Test
($N = 45$ males; 43 females.)

Test	Mean	*SD*	Reliability*
Homonym test			
Males	22.00	6.04	.63
Females	21.14	5.70	.66
Body-Cathexis			
Males	3.43	.337	.78
Females	3.46	.401	.83
Self-Cathexis			
Males	3.43	.393	.88
Females	3.35	.510	.92
Anxiety indicator			
Males	3.09	.524	.72
Females	3.05	.625	.73

* Corrected by the Spearman-Brown formula.

Split-half reliability coefficients were obtained for all of the above scores. These are given in Table 3. Reliability coefficients for the homonym test are about .15 lower than those obtained in previous studies, but are still satisfactory. Reliabilities for the various BC-SC scores are moderately high; they may be regarded as more impressive when it is recalled that subjects displaying the most consistency were removed from the sample.

No important differences between means of the various scores for the two sexes were obtained.

Table 4

Intercorrelations between BC-SC Scores, the Homonym Test, and the Maslow Test
($N = 45$ males, 43 females; except for the Maslow test, for which $N = 46$ males and females.)

Test	Body-Cathexis	Self-Cathexis	Anxiety Indicator
Homonym test			
Males	–.18	–.15	–.37*
Females	–.41**	–.23	–.40**
Maslow test	–.37**	–.52**	–.41**
Self-Cathexis			
Males	.58**		
Females	.66**		

* Significant at the .05 level.
** Significant at the .01 level.

Relationship between body-cathexis and self-cathexis. Intercorrelations between total BC scores and total SC scores are shown in Table 4. The *r* for men is .58, and for women, .56, suggesting that individuals have a moderate tendency to cathect their body to the same degree and in the same direction that they cathect their self; that is, the two kinds of cathexis covary.

Relationships between body-cathexis and the homonym test. An examination of Table 4 reveals no significant correlations between the H test and BC, for the 45 males, or between the H test and SC. There was, however, an r of $-.37$ between the H test and the anxiety-indicator score which, although low, was significant at the .02 level. For the 43 females, on the other hand, correlations significant beyond the .01 level were obtained between H-test scores and BC ($r = -.41$), and H-test scores and the anxiety-indicator score ($r = -.40$). Correlations between homonym scores and SC were not significant. In general, then, there is a low but significant relationship between BC as measured by the scale and anxiety-related BC as measured by the H test.

A concrete case may perhaps serve to indicate more specifically the manner in which the BC Scale and the H test concern anxiety.

A woman with an extremely high H-test score and an extremely low BC score responded to the test items in the following way:

1. Items on the BC Scale rated in category 1 (*Have strong feelings and wish change could somehow be made*) were: facial complexion, appetite, hands, distribution of body hair, nose, fingers, waist, energy level, ears, body build, skin texture, knees, and weight. Items rated category 2 (*Don't like, but can put up with*) were: hips, sleep, sex activities, and posture. None of the 46 body items were rated in category 5 (*Consider myself fortunate*).

2. On the homonym test, responses to meanings pertaining to the body, body processes, or pain and disease are shown below, preceded by the homonym stimulus word in italics: *acid*-sour, *arch*-bend, *back*-broken, *bare*-ugly, *condition*-sick, *crisis*-sick, *enlarged*-grown, *extract*-tooth, *function*-stomach, *gag*-whiskey, *gall*-bile sac, *gas*-dying, *glassy*-dead, *patient*-doctor, *rash*-heat, *red*-hands, *scarlet*-disease, *side*-hurt, *sling*-broken arm, *stain*-blood, *stiff*-corpse, *strip*-New Orleans, *system*-biology, *tablet*-sleeping pill, *tan*-Daytona Beach, *tape*-sprained ankle, *tender*-skin, *trunk*-body, and *waist*-starve.

These item responses give some idea of the importance of the body to this woman and reveal her anxiety concerning it. A consideration of the diverse nature of the processes involved in the BC Scale and the H test as revealed by these items lends significance to the correlation obtained between these two tests.

Sex differences in degree of cathexis. It was found that females cathect their bodies, irrespective of direction, more highly than do males, in that they do not assign as many *3*'s to body items (3 – Have no particular feelings one way or the other). The mean number of *3*'s for all of the 46 body items was 10.76 for men, and 7.86 for women. This difference of 2.90 yielded a critical ratio of 1.92, which approached significance at the .05 level. Consistent with this datum is the greater variability of BC scores among women, as previously shown in Table 3, although this difference in sigmas is not significant.

Relationships between the Maslow test and the BC-SC Scale. The Maslow test was found to correlate $-.37$ with BC, $-.41$ with the anxiety-indicator score, and $-.52$ with SC. These r's are significant at the .01 level. It may be concluded that low cathexis is associated with insecurity to some degree.

Discussion

One of the most significant results is the demonstration that the body and the self tend to be cathected to the same degree. This supports the hypothesis that valuation of the body and the self tend to be commensurate.

Since the persons who obtain a high score on the H test are regarded as being more anxious concerning their bodies than those obtaining a low score, a moderate negative correlation would be anticipated between BC and the H test. This is consistent with the negative r actually obtained for women.

In the case of men, the relative lack of significant relationships between the homonym test and the BC-SC Scale cannot be clearly explained on the basis of available data. A likely but unproved hypothesis is that women are more likely than men to develop anxiety concerning their bodies, because of the social importance of the female body. This hypothesis is consistent with the datum that woman cathect their bodies more highly than men, irrespective of the direction of cathexis, and with the fact that when a BC score based only on those body items most frequently arousing negative feelings in men was employed, a significant r of $-.37$ was obtained for men between this "anxiety-indicator" score and H-test scores.

A final support for the interpretation of BC as a self-related variable is found in the significant correlation between BC and insecurity, as measured by the Maslow test.

Summary and Conclusions

The purpose of the present study was to develop a method for appraising the feelings of an individual towards his body and to ascertain whether or not variables derived from these appraised feelings are significant for personality theory. The latter was determined by testing the hypotheses stated under 2, 3, and 4 below.

A scale for the determination of the degree of cathexis towards various aspects of the body was designed and administered to groups of college males and females, along with a similar scale for aspects of the self, a homonym test of anxiety-related body-cathexis, and the Maslow Test of Psychological Security-Insecurity.

The following conclusions may be drawn:

1. The split-half reliabilities of the two parts of the scale were found to be satisfactory: .81 for body-cathexis, and .90 for self-cathexis.

2. The hypothesis that feelings about the body are commensurate with feelings about the self was supported by significant correlations between the two parts of the scale.

3. The hypothesis that low body-cathexis is associated with anxiety in the form of undue autistic concern with pain, disease, or bodily injury was upheld by the demonstration of significant relationships between low body-cathexis as determined by the scale and by the homonym test.

4. The hypothesis that low body-cathexis is associated with insecurity was sustained by the demonstation of correlation between the former and the Maslow test.

Received January 19, 1953.

References

1. Cronbach, L. J. Response sets and test validity. *Educ. psychol. Measmt,* 1946, **6,** 475-494.
2. Cronbach, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt,* 1950, **10,** 3-31.
3. Maslow, A. H., Hirsh, E., Stein, M., & Honigmann, I. A clinically derived test for measuring psychological security-insecurity. *J. gen. Psychol.,* 1945, **33,** 21-41.
4. Schilder, P. *Psychotherapy.* New York: Norton, 1938.
5. Secord, P. F. The homonym word association test. Unpublished doctoral dissertation, Stanford Univer., 1950.
6. Secord, P. F. Objectification of word association procedures by the use of homonyms: a measure of body-cathexis. *J. Pers.,* 1953, **21,** 479-495.

Journal of Consulting Psychology
Vol. 17, No. 5, 1953

Prediction of Duration of Therapy from The Rorschach Test[1]

Robert G. Gibby, Bernard A. Stotsky,
Veterans Administration Mental Hygiene Clinic
Detroit, Michigan

Daniel R. Miller, and E. Wesley Hiler
University of Michigan

This is the first of two investigations designed to study differences in the Rorschach protocols of patients who remain in therapy and those who do not. The present paper is concerned with the meanings of the scoring categories that differentiate between these groups. A subsequent report will be devoted to the development and validation of a pattern of signs that can be used in the clinic to predict duration of therapy.

In the establishment of valid prognostic signs on any projective instrument, the researcher is beset by numerous difficulties. The patient's behavior in therapy is related not only to his character structure but also to such interacting sources of variance as the therapist's skill, stability in the patient's working hours and his place of residence, and financial arrangements. In connection with the interpretation of the Rorschach protocol, the experimenter must take into account such sources of variance as the setting in which the test is administered, the patient's ego-involvement, and the examiner's character structure [4]. Finally the success of the experiment may depend upon the accuracy of decisions concerning such aspects of the design as the nature of the sample, kinds of controls, and appropriate cutting points in duration of therapy for identifying the different groups. Negative results are indeterminate, while positive results raise more questions concerning their theoretical significance than they answer.

In view of the bewildering complexity of variables, it is not surprising that there are very few studies of prognosis in therapeutic or other situations. However, an increasing number of investigations are being motivated by the fact that clinicians stake their professional reputations upon their abilities to make even more complex analyses than those of the present study. For example, reports may contain predictions of the patients' relationships to the therapists or warnings of the probable dangers of depth therapy. While the interpretations are sometimes incorrect, many psychologists can cite instances of predictions of behavior that were borne out in great detail, far beyond chance, by subsequent events. It is hoped that investigations such as the present one will contribute to an understanding of the processes that are involved in the development of these accurate predictions.

Subjects

Setting. Subjects were selected from among patients at the Veterans Administration Mental Hygiene Clinic, Detroit, Michigan. The work of this clinic is performed by five teams, each consisting of a psychiatrist, psychologist, social worker, and two or three trainees in clinical psychology. To each team are assigned all patients who register on a particular day of the week. Since there is no policy of scheduling any type of problem for registration on a particular day, various pathologies are randomized among the different teams. A tabulation of the incidence of different syndromes in the subjects of the present study revealed no significant differences among the teams.

After the patient registers, he is first inter-

[1] From the Veterans Administration Regional Office, Detroit, Michigan.

viewed by the social worker, who notes his complaint and source of referral, takes a case history, and writes a report attempting to formulate his dynamics. Then the psychiatrist conducts a diagnostic interview. Following preliminary discussion by the team, the case may be referred to the psychologist for diagnostic evaluation. After this procedure, the team holds a planning conference to review the data and decide upon final recommendations. When they decide upon therapy, the patient may be referred to either a psychiatrist, psychologist, or social worker.

General criteria. For purposes of simplifying the analysis of data, a number of criteria were utilized in the selection of the initial pools of subjects. The age range was restricted by eliminating veterans of the first World War. Only males were used. Finally, cases of organic brain damage and disoriented psychotics were rejected in order to maximize cooperation in the administration of tests and insure the possibility that all subjects could have profited from therapy.

Criteria for selection of groups. This investigation involved the comparisons of the protocols of three groups. All members of the first group were receiving psychotherapy at the time data were collected and had been in treatment continuously for at least six months. Because visits to the clinic by different patients varied between one and three times per week, the number of sessions in this group ranged between 27 and 305; the mean is 121.

Group II consisted of subjects who discontinued therapy prior to the sixth session against the therapist's advice. Care was taken not to include individuals who had to terminate treatment because of circumstances beyond their control, such as change in working hours or of residence. The number of therapeutic appointments ranged between one and five, with a mean of 2.7. In order to control for the skill of the therapist, the design required that each contribute an equal number of patients to the first two groups. Sixty-six cases were found who conformed to the general criteria and could be paired in terms of their common therapists. Thus 33 were assigned to each group.

Like the other subjects, members of Group III were processed by their teams and advised to undertake psychotherapy. Unlike the others, they refused treatment. Of the cases in the files who met the criteria, 33 were chosen at random, care being taken not to include those who declined therapy because of situational factors beyond their control.

Therapists. Of necessity, the procedure of selecting patients simultaneously determined the population of therapists. It may help in the interpretation of results to discuss some of the characteristics of this population. There were ten therapists in all, four psychiatrists, four clinical psychologists, and two social workers. In the three professional groups, there were considerable variations in the amounts of experience with adult neurotics and psychotics. The psychiatrists varied between one and five years of practice beyond the residency. Among the psychologists, the least experienced had received a full year of supervised therapy beyond the completion of the doctoral degree; the most experienced had spent half of her twelve years of clinical work in therapeutic practice. The two social workers had two to five years of practice in psychotherapy beyond their formal training.

In the Detroit clinic, cases are presented regularly in staff conferences conducted by psychiatric consultants, who are all psychoanalysts. The considerable influence of these consultants is manifested by the psychoanalyt-

Table 1

Characteristics of Population of Patients

Variable	I. Six months or longer	II. Five sessions or less	III. Refused therapy
N	33	33	33
Mean age	31	28	30
Education: mean grade	10.8	10.5	9.4
No. of therapists	10	10	----
White	31	30	24
Negro	2	3	9
Referred by medical clinic	20	21	22
Referred by vocational rehabilitation	1	3	4
Self-referred	12	9	7
Diagnosed neurotic	27	22	18
Diagnosed schizophrenic reaction	4	8	8
Diagnoses deferred	2	3	7

Table 2
Differences in Absolute Scores of Groups

Rorschach Score	Mean of Group I	Mean of Group II	Mean of Group III	*SD* of Group I	*SD* of Group II	*SD* of Group III	CR I-II	CR I-III	CR II-III
R	31.5	20.2	18.2	17.5	9.7	8.3	3.2**	3.9**	1.0
W	6.6	5.1	6.4	4.9	3.7	3.3	1.4	.2	1.5
D	20.8	12.7	10.6	9.3	7.5	3.7	3.9**	6.0**	1.4
Dd	4.1	2.4	1.1	5.0	2.8	2.0	1.7†	3.2**	2.2*
S	2.1	1.1	1.7	2.3	1.3	1.8	2.2*	.8	1.5
F	16.1	10.9	9.7	8.6	6.6	5.3	2.9**	3.6**	.8
F+	10.0	7.8	7.9	5.5	4.3	4.6	1.8†	1.9†	.1
M	2.5	1.7	1.7	3.3	1.8	1.8	1.3	1.3	----
FM	2.7	2.4	2.0	2.3	2.0	1.8	.6	1.4	.9
m	.9	.2	.2	1.5	.3	.4	2.7**	2.7**	----
m+add. *m*	2.3	.5	.5	3.4	.9	.4	3.0**	3.0**	----
b	.7	.1	.1	1.1	.5	.3	3.9**	3.0**	----
K	.5	.1	.3	.8	.4	.6	2.5*	1.2	1.5
FK	1.6	.6	.5	.8	.8	1.1	5.0**	4.6**	.4
Fc	2.3	1.1	1.3	2.9	1.1	1.4	2.2*	1.8†	.7
c	.5	.2	.1	.9	.4	.3	1.8†	2.4*	1.0
C′	.4	.4	.8	.7	.7	1.0	----	2.0*	2.0*
Sh	5.6	2.5	3.2	5.5	2.1	2.8	3.0**	2.2*	1.2
FSh	4.2	2.2	2.8	4.1	2.0	2.4	2.5*	1.7†	1.1
ShF	1.4	.3	.4	2.0	.5	.7	3.1**	2.8**	.7
FC	1.7	1.4	.7	1.6	1.4	.9	.5	3.1**	2.5*
CF	1.9	1.1	.8	1.9	1.4	1.2	2.0	2.3*	1.0
C	.3	----	.1	.5	----	.4	3.0**	1.8†	1.4
Sum C	3.9	2.5	1.6	3.1	2.1	1.6	2.1*	3.8**	2.0*
A	12.5	10.0	9.2	5.0	6.4	3.9	1.7†	3.0**	.6
H	5.6	3.3	3.2	5.0	2.7	3.0	2.3*	2.4*	.1
P	6.3	5.0	6.0	2.3	1.4	2.2	2.8**	.5	2.2*
No. of Cont. Categories	8.1	5.5	5.4	3.3	1.6	2.4	4.0*	3.8**	.2
An	2.8	1.5	.9	2.5	1.6	1.1	2.5*	3.9**	1.7†
Sex & Anal.	1.2	.1	.0	3.1	.3	.2	2.0*	2.0*	----

† Significant at the 10% level or less.
* Significant at the 5% level or less.
** Significant at the 1% level or less.

ically oriented therapy which is practiced by the entire staff.

The fact that each therapist contributed some cases to the first group guaranteed that he was able to maintain some relationships with patients for at least six months. It might be anticipated that some therapists may have been excluded from the study because they were skilled enough to hold all of their patients and therefore could contribute none to the second group. Actually, this was not the case. All were treating some individuals who, because of their primitive defenses or the fear of losing their pensions, could probably not have been retained in therapy by even the most able of practitioners.

Controls. Table 1 provides a summary of variables which, if they were not controlled, might have contributed to significant differences among the three groups. Of all the differences, only the one involving the proportions of Negroes is significant at the 5 per cent level of probability. Although the total number of Negroes is quite small, the percentage in the third group is significantly higher than that of the first. However, an independent analysis revealed no significant relationships between race membership within the groups and differentiating Rorschach variables.

Results and Discussion

In order to maximize reliability, all protocols were rescored by one individual according to Beck's criteria for location, *F*+, *P*, and con-

Table 3

Differences in Percentage Scores of Groups

Rorschach Score	Mean of Group I	Mean of Group II	Mean of Group III	*SD* of Group I	*SD* of Group II	*SD* of Group III	CR I-II	CR I-III	CR II-III
W	24	30	42	16	20	22	1.3	3.8**	2.4*
D	65	60	54	15	15	19	1.4	2.6**	1.4
Dd	10	11	4	9	11	7	.4	3.0**	3.2**
S	7	5	9	6	6	10	1.3	1.2	2.4*
F+	82	89	90	16	15	15	1.8†	2.1*	.3
F	53	54	54	17	18	20	.2	.2	----
M	8	9	9	10	8	9	.5	.5	----
FM	9	12	11	6	8	12	1.7†	.9	.4
m	2	1	1	3	2	2	1.6	1.6	----
m+add. *m*	6	2	2	6	4	2	3.1**	3.6**	----
k	2	1	1	4	2	2	1.3	1.3	----
K	1	1	2	2	2	3	----	1.6	1.6
FK	3	3	2	5	4	4	----	.9	1.0
Fc	7	6	7	7	5	7	.7	----	.7
c	1	1	1	2	2	4	----	----	----
C'	1	3	3	3	4	4	2.2*	2.2*	----
Sh	17	12	16	10	10	11	2.0*	.4	1.6
FSh	13	11	13	8	10	10	.9	----	.8
ShF	4	1	3	5	3	5	3.0**	.8	2.0*
FC	6	6	4	3	6	6	----	1.7†	1.4
CF	6	5	4	4	6	6	.8	1.6	1.4
C	1	----	.5	2	----	2	2.9**	1.1	1.5
Sum C	12	11	9	7	8	8	.5	1.6	1.0
8, 9, 10	34	36	28	7	14	9	.7	2.9**	2.8**
A	43	54	55	12	20	16	2.7**	3.4**	.2
H	16	16	16	11	13	12	----	----	----
P	25	30	37	12	15	17	1.5	3.3**	1.8†
No. of Cont. Categories	29	31	31	10	6	9	1.0	.9	----
An	9	6	6	7	6	8	1.7†	1.9†	----
Sex & Anal.	3	1	----	7	3	1	1.4	2.5*	1.4

† Significant at the 10% level or less.
* Significant at the 5% level or less.
** Significant at the 1% level or less.

tent, and Klopfer's criteria for the determinants. Using the same system, a second psychologist then independently scored 20 randomly selected records of the first group and 15 of the second. The reliabilities of the scoring categories for the first group ranged from .79 to 1.00, with a mean of .95. For Group II, the correlations varied between .41 and 1.00, and the mean is .87. Only one coefficient was below .80.

Distributions of the frequencies of both absolute scores and their percentages were then prepared for each of the three experimental groups. Means and sigmas were calculated and the significances of differences among groups tested in terms of critical ratios. Table 2 contains an analysis of differences among the three groups on each of the scoring categories. While clinicians do not customarily convert all the data into percentage scores before making their interpretations, this has been done in the present study in order to determine the differences among the groups that remain significant when the number of responses is controlled. The data are contained in Table 3.

Motivation for therapy and resistance. One interpretation of these data is suggested by Calden's [1] comparison of the Group Rorschach protocols of subjects with high and low ego-involvement in the test. The ego-involved subjects, like Group I in the present study, give many responses and obtain high scores in *D* (which usually has a high correlation with *R*), *m*, diffusion and shading responses and color. Those low in ego-involvement are similar to Groups II and III in their stereotypy as indicated by the large proportion of animal responses. In Calden's opinion, the tendencies of the poorly involved subjects to favor the easily located animal responses and to avoid color, shading, and movement reveal a suppressive control that bans spontaneity. If this interpretation is also valid for the subjects of the present study who do not remain in therapy, it raises the possibility that they may wish to avoid changing their behavioral patterns. Such a goal might be motivated by diverse factors such as satisfaction with their current personalities, fear that change may lead to psychosis, or the desire to preserve the symptoms that justify their pensions. Regardless of the underlying motives, the protocols of these patients suggest that they react to the projective test in a constricted and unrevealing manner, and later leave therapy as soon as the possibility arises that they may become emotionally involved.

This explanation is further supported by Schachtel's clinical observations of resistances to the Rorschach test. These may take the form of a fear of being "found out," a defiance privately phrased as: "You are not going to get anywhere with me," or an apparent listlessness that serves as "the cover for an unconscious quite active resistance" [7, p. 432]. In his efforts to achieve a noncommittal record the patient often limits his total productivity as well as the human and human movements responses. The record generally becomes stereotyped in content because of the high proportion of animal and popular responses. In the present study, all the scoring categories listed by Schachtel except human movement, the trend of which conforms to his observations, differentiate significantly between the first group and either the second or third in the predicted directions.

Other findings tending to support the interpretation of resistance are the small number of content categories in the second and third groups and their paucity of sexual and anal references. Finally a type of censorship by groups who do not stay in therapy is indicated by their large $F+$ percentages. According to Rorschach, a high proportion of $F+$ is a sign that the subjects are "bringing to bear all their attention and faculties of self-criticism . . . though the answers are extraordinarily stereotyped, showing a poor range of variation" [7, p. 57].

In summary, the Rorschach records of the poor therapeutic risks are characterized by limited productivity, few content categories, stereotypy, a paucity of references to sexual and anal topics, little color, and preferences for good form responses. If these scoring categories are interpreted in the conventional manner, they indicate that these patients are probably overly deliberate in therapy, do not talk much, tend to stick to "safe" subjects when they do talk, and censor their emotional reactions.

Conscious anxiety. An additional interpretation of the differential scoring symbols is suggested by the types of initial complaints. These may be divided into three categories, psychological, somatic, and a combination of the two. Examples of the psychological complaints are phobia, obsession, compulsion, anxiety, amnesia, worry, irritability, nightmare, and depression. Among the somatic complaints are gastric pains, diarrhea, constipation, hysterical blindness, headaches, excessive perspiration, anorexia, breathlessness, and hives. Table 4 lists the frequencies of the different types of symptoms in the three groups.

Table 4

Initial Complaints

	I. Six months or longer	II. Five sessions or less	III. Refused therapy
Somatic	5	16	13
Psychological	9	9	9
Combination	18	8	10
Undetermined	1	----	1

The chi square of 10.6 is significant at the 5 per cent level of probability. While a majority of Group I complain of a combination of psychological and somatic symptoms, the largest proportions of those in the other groups tend to seek help solely for somatic symptoms. The differences between the first and second groups are significant at the 5 per cent level. A probability of 6 per cent is obtained for the comparison of the first and third. Since the differences between the second and third groups could have been obtained by chance, they may be considered homogeneous with respect to type of complaint.

Fenichel [2] notes that somatic symptoms are alternatives to anxiety as derivatives of forbidden impulses. Usually, the more the patient complains of somatic symptoms, the less concerned he is with the problems that create them. From these observations, it was deduced that patients who tend to somatize to the exclusion of psychological problems should manifest fewer signs of anxiety on their Rorschach protocols than others. This tends to be the case. Of the indicators of anxiety listed by Klopfer [3], the two groups who do not remain in therapy have significantly less *FK, k, Fc,* and *m* than the group that remains in therapy. Not significant but in the predicted direction is the paucity of *K* in the records of the somatizing groups.

A Comparison of Two Experiments

Judged by the number of significant differences, Group I, consisting of those who stayed in therapy at least six months, is very different from the subjects who refused therapy or did not continue after the fifth interview. On the other hand, in terms of the signs of cooperativeness and ability to face anxiety there seems to be little difference in the motivations of patients in this clinic who see a therapist less than six sessions and those who refuse therapy outright. This problem of determining an appropriate cutting point suggests one explanation of the discrepancy between the present results and those of another investigation of duration in therapy by Rogers, Knauss, and Hammond [5]. These investigators report a lack of significant differences between the protocols of subjects who stayed in therapy and those who left early. In that design, there were two groups: one had four or less therapeutic sessions and the other had five or more. Reasons for the choice of this cutting point are not presented. It is conceivable that there were no significant differences in the motivations for therapy or the defenses of those who had four therapeutic sessions and those who had five, six, seven, or even twenty appointments. This problem was avoided in the present experiment which utilized extreme groups, those with five sessions or less and those who stayed in therapy at least six months. Other modifications of the design by Rogers, Knauss, and Hammond were the controls of factors that could reduce the significance of relationships between Rorschach signs and duration of therapy. These controls included sex, source of referral, disorientation, dropping of therapy because of reasons beyond the control of the patient, and therapist's skill. Only further studies of individual variables can determine which aspects of the designs of the two investigations contributed significantly to the differential results.

Summary

This is a study of the Rorschach protocols of three groups: one remained in therapy for at least six months; another terminated treatment in less than six sessions against the therapist's advice; and a third refused therapy when it was recommended.

Significant differences between the first and the other two groups on *R*, $A\%$, *H*, and $F+\%$, are interpreted as indicating a greater motivation for therapy and resultant cooperativeness on the part of the former. The stereotypy and lack of productivity of the other groups seem to indicate resistance to becoming involved in the testing and therapeutic procedures. Significant differences in the scoring categories total *C, FK, k, Fc,* and *m* are interpreted as revealing a greater awareness of anxiety on the part of the first group in contrast to the suppressive reaction of the two who do not remain in therapy.

The present design is compared with that of Rogers, Knauss, and Hammond in order to account for differences in results. The blind empiricism inherent in investigations of the prediction of behavior in therapy from projective tests is stressed, and an effort is made to summarize some of the principal sources of variance that require more detailed investigation.

Received January 8, 1953.

References

1. Calden, G. The relationship of varied test definitions and degrees of ego involvement to Rorschach test performance. Unpublished doctor's dissertation, Univer. of Michigan, 1951.
2. Fenichel, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.

3. Klopfer, B., & Kelley, D. McG. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1942.
4. Miller, D. R. Prediction of behavior by means of the Rorschach test. *J. abnorm. soc. Psychol.,* 1953, **48**, 367-375.
5. Rogers, L. S., Knauss, J., & Hammond, K. R. Predicting continuation in therapy by means of the Rorschach test. *J. consult. Psychol.,* 1951, **15**, 368–371.
6. Rorschach, H. *Psychodiagnostics.* Berne: Hans Huber, 1942.
7. Schachtel, E. G. Subjective definitions of the Rorschach test situation and their effect on test performance. *Psychiatry,* 1945, **8**, 419–448.

Journal of Consulting Psychology
Vol. 17, No. 5, 1953

The Rorschach as a Measure of Real-Life Stress

David Berger
Michigan State College [1]

Despite the present popularity of the Rorschach test as a clinical tool there is no adequate body of validating data to confirm the numerous assumptions underlying the test's use. The growing concern with this deficiency has begun to provide impetus for controlled experimentation which should result in a better understanding of the instrument.

One of the major problems in the evaluation of the test has been to appraise its efficacy in measuring anxiety. Much of the work purporting to treat this problem has been conducted under conditions foreign to the clinical application of the test. The list of these efforts includes the use of the group Rorschach technique as employed by Smith and George [7], and the attempts by Williams [8], Eichler [2], and Levine, Grassi, and Gerson [4] to employ artificial stress to induce anxiety. The results obtained in these studies, while suggesting that stress influences Rorschach performance, are not necessarily pertinent to the usual patient-examiner relationship. In a more recent experiment by Klatskin [3] an acute, real-life, temporary stress situation was selected to measure the influence of anxiety on the test results. In contrast to the previous studies her experimental setting more closely approximated the testing situation encountered in clinical practice. She was able to substantiate 22 of 26 Rorschach indices commonly used in predicting adjustment, and concluded that the Rorschach was sensitive to real-life stress. On the other hand the literature contains conflicting evidence suggesting that the test has questionable validity in distinguishing overtly anxious groups from normals. For example, Young and Higginbotham [10], as well as Wischner, Rotter, and Gillman [9], reported no significant correspondence between emotional disturbance and certain "anxiety indicators." These studies have created a general feeling of scepticism about the test. Further investigation appears to be necessary to resolve this controversy.

Problem

The present study was designed to ascertain the influence of a real-life stress on the Rorschach test. It was concerned with the response to stress, as well as with the recovery from stress over a period of time. Specifically, two questions were asked: Is the Rorschach sensitive to a real-life stress situation, and if so, in what respect does the Rorschach reflect the stress?

Procedure

Sample. Two matched groups each consisting of 40 persons were tested and retested six weeks later with the Rorschach.[2] The experimental group was selected from a population of persons on the day of their first admission to a tuberculosis hospital or a sanatorium.[3] Patients who were in acute medical stress on admission were not seen. A range of 18 through 46 years was imposed to reduce the influence of age factors on Rorschach performance [7]. The control group was selected exclusively from the sanatorium population. Only patients whose hospitalization extended at least six continuous months were used. They were matched individually with the experimental subjects on the basis of age, sex, race, marital status, education, and diagnostic classification [6]. The results of the matching are presented in Table 1.

The tests were administered and scored according to the method developed by Beck [1].

---

[1] The writer is now staff psychologist at Fort Custer, VA Hospital, Fort Custer, Michigan.

[2] Other measures were utilized and will be reported in a subsequent article.

[3] Herman Kiefer Hospital and Wm. Maybury Sanatorium are Wayne County tuberculosis facilities serving the City of Detroit and its environs.

Table 1

Summated Characteristics of Experimental and Control Samples

Population variables	Experimental sample	Control sample
Sex		
Male	21	21
Female	19	19
Race		
White	38	38
Negro	2	2
Marital status		
Married	31	32
Single	6	6
Separated	2	2
Widowed	1	0
Mean age	31.10	30.98
Mean education (grade)	10.23	10.30
Extent of illness		
Minimal findings	4	3
Moderate findings	26	26
Far advanced findings	10	11

Both groups were given the impression that the testing procedure was an integral phase of the treatment program, but participation on the part of the patient was optional. After completion of the initial test it was explained to each patient that he would be seen again in a number of weeks.

Stress situation. It was assumed that most patients entering a tuberculosis hospital perceive their admission as a threatening experience. It seemed reasonable to consider that such a situation would generate anxiety and should conceivably afford an opportunity to measure anxiety in human beings outside the experimental laboratory.

A six-week interval between test and retest was used with both groups. It was hoped that this interval would allow the newly arrived patient ample time to make some adjustment to the hospital environment. The expectation was that his readjustment would be reflected in the change from test to retest. The control group was employed to control the questionable reliability of the test. Because of the relatively stable environment of the sanatorium it was felt that the interval between tests would not contribute any significant variance to the retest performance of the control population.

Statistical treatment. The Rorschach results were handled in terms of the global configuration of the performance, as well as by analyzing the various atomistic changes. In dealing with the global Rorschach, two judges were asked to select the more anxious record from the two records of each experimental subject. The atomistic procedure was carried out by the comparison of differences between the test results from test to retest in the two groups. The difference in scores between test and retest was computed for each subject. Each experimental subject's difference score was compared with the difference score earned by the matched subject in the control group. The distribution of differences between difference scores was utilized to yield t values. McNemar [5] has discussed and justified this method for treating the relative shifts in two groups over a period of time. The difference between the shifts in the two groups was accepted as a behavioral response to stress.

The atomistic analysis of the Rorschach protocols was based on the absolute number in each scoring category. The traditional objection to a raw-score procedure in Rorschach experimentation is that when two populations produce deviant response totals, any difference in the scoring variables is likely to be a function of the differences in the response totals. Many experimenters have attempted to satisfy this criticism by converting into percentages and other statistical derivatives. It is questionable whether these corrections are sound or really necessary. The view taken in this research was that response total is in reality the dependent variable, mirroring dynamic changes in personality as demonstrated in the approach, determinant, and content scoring factors. The evidence points to a differential change in the scoring variables attending a change in response totals. If the various scores were the dependent factors that they have ordinarily been conceded to be, it should follow that they vary directly with changes in response total. The present data, in fact, show many instances in which an increased total response is accompanied by a decrease in certain other scores. It seemed more parsimonious to view response total as a summary statistic dependent upon the remainder of the scoring profile.

Results

Configurational findings. The problem of the reliability of the judgments made on the total test performance by the two raters was handled by chi square. The two judges were in agreement in their evaluation on 33 of 40 pairs of records, that is, they selected the same record as being the more anxious of the two records obtained from each experimental subject. This result is significantly better than chance at the 1 per cent level.

The validity of the judgments was similarly subjected to a chi-square test. One judge made 35 "correct" judgments, i. e., selected the first test, out of 40 judgments, while the second judge was able to identify "correctly" 36 of the 40 pairs of records. The chi square for each judge was significantly better than chance at the 1 per cent level.

Atomistic findings. A comparison of the shifts of the two samples described from test to retest for the various Rorschach scores is summarized in Table 2. There were statistically significant shifts in 12 of the 16 variables assayed. These shifts were presumably the result of the anxiety associated with the admission situation. In general the direction of the shifts as well as the kind of change noted is consistent with the "anxiety indicators" commonly used in clinical practice (Table 2).

Stress was demonstrated to influence the Rorschach in this study by: (*a*) reducing responsivity (*R*), (*b*) enhancing reality testing ($F+\%$), (*c*) decreasing the breadth of interest (number of content categories), (*d*) increasing reaction time ($T/1R$), (*e*) increasing the number of whole comprehensive responses (*W*), (*f*) decreasing the number of common detail responses (*D*), (*g*) decreasing the amount of inner living and creativity (*M*), (*h*) decreasing the emotional participation with the environment (Sum *C*), (*i*) increasing dysphoric feelings, and feelings of inferiority (Sum *Y* and Sum *V*), (*j*) increasing stereotypy and evasiveness ($A\%$), (*k*) reducing the ability to think in terms of group standards (*P*), and (*l*) intensifying the need to reject environmental stimulation (rejections).

Table 2

Comparison of Change from Test to Retest between Experimental and Control Groups

Rorschach score	Mean change from test to retest in exp. group	Mean change from test to retest in control group	Difference between mean changes	σ_d	*p*
R	7.22*	0.42*	6.80	13.19	.01
F plus %	–2.75	7.42	–10.18	17.74	.001
# Content Categories	1.98	–0.18	2.15	3.10	.001
T/lR	–8.85	–1.36	–7.49	10.33	.001
A%	–3.58	2.05	–5.62	14.91	.05
H Hd	2.65	0.80	1.85	6.25	N.S.
W	–2.28	1.15	–3.42	4.19	.001
D	8.88	–0.52	9.40	10.30	.001
Dd	0.60	0.48	0.12	4.18	N.S.
M	1.80	–0.02	1.82	3.16	.001
Sum *C*	1.59	–1.12	2.71	3.13	.001
Sum *Y*	–1.08	0.49	–1.56	1.52	.001
Sum *V*	–0.14	0.09	–0.22	0.68	.05
P	2.58	–0.50	3.07	2.18	.001
S	–0.20	–0.70	0.50	1.97	N.S.
Card Reject	–15†	–2†	(not calculated)		

Note—Significance was determined by two-tailed test.
* A positive score means an increase from first to second test.
† Total difference in number of rejections.

Discussion

The response of the experimental population seems to have been influenced markedly by the stress of entering a tuberculosis institution. With two exceptions (*W* and $F+\%$) all the shifts in Rorschach performance are consistent with the alleged indices of anxiety. The present study is at variance with previous findings in that stress appeared to facilitate the perception of whole responses, and increased the need to adhere closer to reality. The Rorschach records of the experimental group in the first test were then analyzed to determine the type of *W* response elicited. A large preponderance was found to be ready-made, easy wholes. They probably represent the most elementary type of response, involving a minimum of effort for the subjects under stress.

The present findings support the contention that the Rorschach test is a useful instrument for measuring anxiety in the type of clinical situation studied here. The conflicting view that the Rorschach has questionable validity for measuring anxiety may stem from a different experimental emphasis, utilizing artificial laboratory stress or diagnostically anxious subjects. The present study, as well as the one reported by Klatskin [3], suggests that real-life stress is an effective medium for measuring the effect of anxiety on Rorschach performance. It seems reasonable to presume that a real-life stress condition results in more pervasive discomfort than the distress aroused by a noxious laboratory stimulus.

The effectiveness of the Rorschach in this investigation contributes to the growing realization that it is an instrument sensitive to fluctuations in the life situation of the subject. This realization may serve to emphasize the pitfalls involved in interpreting the test clinically without a thorough awareness of the current environmental pressures impinging on the individual.

Summary

Two matched groups of tuberculous patients were tested with the Rorschach and retested six weeks later. One group was seen initially on the day of their admission to a tuberculosis treatment facility. The second group was seen after they had been hospitalized for at least six months. The major hypothesis was that the initial admission situation would provoke a stress reaction recognizable on the Rorschach test.

Other than for minor discrepancies, the results obtained were consistent with anxiety signs used in clinical practice. It was also demonstrated that the Rorschach could be used as a measure of total personality adjustment in a stress situation. Certain cautions were discussed in regard to the need to evaluate the immedediate life situation in using the Rorschach in personality analysis.

Received January 15, 1953.

References

1. Beck, S. J. *Rorschach's test.* I. *Basic processes.* New York: Grune & Stratton, 1944.
2. Eichler, R. M. Experimental stress and alleged Rorschach indices of anxiety. *J. abnorm. soc. Psychol.*, 1951, **46**, 344-355.
3. Klatskin, Ethelyn H. An analysis of the effect of the test situation upon the Rorschach record: Formal scoring characteristics. *J. proj. Tech.*, 1952, **16**, 193-199.
4. Levine, K. N., Grassi, J. R., & Gerson, M. J. Hypnotically induced mood changes in the verbal and graphic Rorschach. A case study. *Rorschach Res. Exch.*, 1943, **7**, 130-144.
5. McNemar, Q. *Psychological statistics.* New York: Wiley, 1949.
6. National Tuberculosis Association. *Diagnostic standards and classification of tuberculosis.* New York, 1950.
7. Smith, S., & George, C. E. Rorschach factors related to experimental stress. *J. consult. Psychol.*, 1951, **15**, 190-195.
8. Williams, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1947, **11**, 20-29.
9. Wischner, G. J., Rotter, J. B., & Gillman, R. D. Projective techniques. In S. W. Bijou (Ed.), *The psychological program in A.A.F. Convalescent Hospitals.* A.A.F. Aviat. Psychol. Progr. Res. Rep. No. 15. Washington: U. S. Government Printing Office, 1947. Pp. 144-157.
10. Young, R. A., & Higginbotham, S. A. Behavior checks on Rorschach method. *Amer. J. Orthopsychiat.*, 1942, **12**, 87-94.

Journal of Consulting Psychology
Vol. 17, No. 5, 1953

The Relation of Rorschach Free Association to Inquiry[1]

Robert G. Gibby and Bernard A. Stotsky
Veterans Administration Mental Hygiene Clinic
Detroit, Michigan

Basic to the use of the Rorschach technique in present-day research is the assumption that the psychogram is determined by the personality of the subject. In the final analysis the value of the Rorschach as a scientific instrument is dependent upon the validity of the objective indices derived from the test scores. Recently a number of studies have focused attention on factors other than the personality of the subject which determine his Rorschach psychogram. The presence of two major sources of such influences has been demonstrated in previously published studies. These are the "set" of the subject [3, 6, 7, 8] and the stimulus value of the examiner [2, 3, 4, 5, 9]. A major implication of these studies is that obtained scores on the Rorschach determinants cannot be regarded as corresponding solely to the various areas of the subject's personality structure and its functioning.

In order to investigate further the influence of the examiner on the Rorschach psychogram, Gibby studied the changes which occurred from the free association to the final response following the inquiry.

Let us at this point briefly review the general technique of Rorschach administration. It may be considered as being composed of two dependent but nevertheless two clearly differentiated processes. In the first, the free association, the examiner plays a comparatively passive role, with the subject being requested to free associate to the Rorschach plates as they are presented to him. In the second, the inquiry, the role of the examiner is much more active, and as the plates are re-presented he conducts an inquiry designed to "clarify" the previously given free associations of the subject. The conclusions of the previously mentioned research on the Rorschach inquiry have raised important questions which bear on the relationship of the free association to the inquiry. These ultimately have a bearing upon the presumed validity of the entire Rorschach technique. It is the purpose of the present study to investigate certain aspects of the relationships that exist between the Rorschach free association and inquiry.

There are three major areas of interest in the present research. These are: (*a*) the similarity of the free-association record to that following the inquiry, (*b*) the comparative effect of the stimulus value of the examiner on the free association and the inquiry, and (*c*) the personality characteristics of the individuals who show the greatest and least amount of change from free association to inquiry.

It is felt that the questions asked in connection with these major areas raise issues crucial to the utilization of the Rorschach technique in contemporary research. For example, does the present method of scoring a Rorschach protocol, based as it is upon the combined free association and inquiry, yield the most unbiased picture of the subject's personality? Again, to what extent are the Rorschach scores and indices, used in many studies to test various hypotheses, contaminated by factors characteristic not of the subject's personality but rather more of the examiner's? It is hoped that the present study will enable the Rorschach technique to be more effectively utilized in assessing the personality of the subject, and equally important, lead to a reconsideration of methods of analysis of Rorschach data in research.

The study will be concerned with the following specific questions:

---

[1] From the VA Regional Office, Detroit, Michigan.

1. Are there significant differences between the Rorschach determinants produced in the free association and those following the inquiry?

2. Do clinicians differ among themselves in the Rorschach determinants produced in the free association?

3. Is the effect of the stimulus value of the clinician on the Rorschach determinants produced during the free association greater or less than that on the inquiry?

4. Are there any differences in free-association Rorschach determinants between those individuals who change the most and those who change the least from free association to inquiry?

Experimental Design

In order to investigate these hypotheses, an experimental group of 240 patients at the Veterans Administration Mental Hygiene Clinic, Detroit, Michigan was formed. Each of these subjects was assigned to a Rorschach examiner depending upon the day on which he first came to the clinic. No other basis of assignment was utilized. Since the clinic has no policy of scheduling any particular type of patient on any certain day, it was felt that this procedure could be regarded as a random assignment of subjects to an examiner. Twelve examiners were employed, each testing twenty subjects. All examiners had received formal training and supervised experience in the administration of the Rorschach, and it was felt that from this point of view all were equally proficient. They were all either Veterans Administration clinical psychology staff members or advanced trainees assigned to the Detroit Mental Hygiene Clinic. The Rorschachs were given routinely, the inquiry being structured by the examiner in his usual manner. None of the examiners had any knowledge or awareness that the present study was to be conducted.

The Rorschach determinants investigated were: *F, M, FC, CF, C, FY, YF, Y,* and *V*. Beck's [1] scoring techniques were in general employed for scoring these determinants. The free-association responses were first scored for each record. Next, the combined free-association and inquiry responses were scored. Each response was thus scored twice, each scoring being done independently. Reliability of the scoring was then determined. For this purpose 27 records were randomly selected, and they were then rescored by a second person who had been previously trained in the scoring techniques utilized in the study. Criteria for scoring of responses were determined in advance. For example, it was decided that all responses such as "humans pointing" should be scored as *M*.

In order to control the variability possibly introduced by *R*, the number of responses for each determinant in the free-association scoring of each record was converted to a percentage. This was done by dividing the number of responses for each of the determinants by the total number of responses. The same procedure was followed for the inquiry scorings. Thus two percentages were calculated for each determinant in each record — the first being the percentage of the determinant in the free association and the second being the percentage of that determinant following the inquiry. In order to treat the data more adequately statistically, these were then translated into arc-sine units. It was thus possible to analyze the data in terms of both the absolute and the percentage scores. The appropriate statistical manipulations were then performed in light of the particular hypothesis to be tested.

Results

The first concern was reliability of scoring. Pearson product-moment correlations between two scorers were computed for each deter-

Table 1
Comparison of Means of Rorschach Determinants in Free Association and Inquiry †

Determinant	Absolute score		Arc-sine	r Between free assn. and inquiry
	Mean diff.	CR	CR	
F	–4.79	17.10**	26.2**	.97
M	.30	4.29**	7.0**	.97
FY	2.14	13.37**	21.4**	.21
YF	.29	9.67**	7.3**	.54
Y	.02	2.00**	1.7	.71
FC	.96	12.00**	14.8**	.58
CF	.90	11.25**	13.1**	.63
C	------	--------	------	.92
V	.19	4.75**	5.1**	.61
R	------	--------	------	1.00

† $N = 240$.

** Significant beyond the 1% level.

minant. This was done for both free association and inquiry. The mean reliability for free-association scoring of the experimental determinants was .91, and for inquiry .92.

In order to test the first hypothesis in regard to differences between free association and inquiry, critical ratios were calculated on both the difference in arc-sines and the difference in absolute scores for each determinant from free association to inquiry. These data are presented in Table 1. It may be summarized from this table that:

1. The following Rorschach determinants change significantly from free association to the inquiry: *FY, YF, FC, CF, F, M,* and *V*. No significant change occurs on *Y* or *C*. These changes occur when the determinants are compared on both percentage and absolute scores.

2. The determinants may be differentially ordered as to the amount of change shown from free association to inquiry. *F* shows the greatest change, followed by *FY*. Next come *FC* and *CF*, followed in order by *M, YF, V,* and *Y*. *F* decreases from free association to inquiry while all the other determinants show an increase.

3. The correlations between the determinants from free association to inquiry also may be differentially ordered. Those showing a high degree of relationship are *F, M,* and *C*. Those showing a moderate relationship are *Y, CF, V, FC,* and *YF*. There is comparatively very little correlation between the number of *FY* responses on the free association and those following inquiry.

In order to explore the second and third hypotheses in regard to the variance attributable to the stimulus value of the examiner, an analysis of variance was done between the distributions of arc-sines for each of the Rorschach determinants on the free association. An earlier study [5] reported the differences between examiners following inquiry. The results of these analyses are:

1. On the free association there are no significant differences between examiners for any of the Rorschach determinants.

2. On the inquiry there are significant differences beyond the 5% level in the cases of the *FY* and *Y* determinants.

A prior study of Gibby [4] showed that examiners differed significantly among themselves in changes from free association to inquiry. That there is no difference between examiners on the free association and that there is such a difference on the inquiry is demonstrated in the present study. Therefore, the free association appears to be free of examiner influence and whatever examiner influence occurs on the Rorschach is attributable to the inquiry.

The third question was concerned with the personality characteristics of those individuals who changed the most from free association to inquiry as opposed to those who changed the least. In order to investigate this, distributions were first made of the changes by subjects in *F* from free association to inquiry in arc-sine units. Changes in *F* reflect the summation of the changes in all other variables since the change in every instance was from *F* in the free association to another determinant in the inquiry. The data were then analyzed in two ways: (*a*) with regard to examiner, and (*b*) without regard to the examiner.

It has been shown that examiners differ among themselves in the amount of change produced from free association to inquiry. It is possible that those examiners producing the least change had within their group of subjects individuals who were capable of change, but due to the influence of the examiner showed comparatively little change. Therefore the six examiners producing the greatest amount of change were utilized. For each examiner the five individuals showing the greatest amount of change and the five individuals showing the least amount of change in *F* from free association to inquiry were selected and combined to form two groups. The critical ratios between these were calculated for each of the Rorschach determinants in the free association. It was found that those individuals who changed the most from free association to inquiry showed a significantly higher number of *F* and a significantly lower number of *M* and *FC* responses in the free association.

In addition a new sample was chosen without respect to examiner. The 40 individuals showing the most change were compared with the 40 showing the least change. Here also those individuals changing the most from free association to inquiry showed a significantly higher number of *F* and a significantly lower number of *M* and *FC* responses in the free association. These data are presented in Table 2.

Table 2

Comparison of Free-Association Determinants of Groups Changing Most and Least from Free Association to Inquiry

Variable	Critical ratio † Examiner uncontrolled (N = 80)	Examiner controlled (N = 60)
F	2.5*	5.6*
M	4.0*	4.7*
FY	.6	.5
YF	1.8	1.0
Y	----	----
FC	2.6*	3.1*
CF	1.1	1.6
C	.5	1.9
V	.3	1.2

* Significant beyond the 5% level.

† Critical ratios are based upon arc-sine differences.

Discussion

Like other psychological techniques the Rorschach is utilized in a dual capacity. In one it plays the role of a clinical instrument, being used to explore the various aspects of the personality of an individual patient. Here the contribution of the technique may be considered as being directly dependent upon the skill of the clinician, and the administration and interpretation of the Rorschach become in practice highly subjective processes. The inquiry itself varies both with the examiner and with the patient and may assume, for example, a directive or nondirective character depending upon the problem which the clinician is called upon to solve. Many clinicians say when this possibility is discussed with them, "Yes, I pushed hard here for shading (or movement or color) but I recognized this and took it into account when I interpreted the record." This is perhaps as it should be since the clinician, in order to deal with his exceedingly complex referral problems, must call upon all of his individual skills and clinical insights. Although it is beyond the scope of the present study, the question might well be raised as to how many of these insights are valid. How many rest upon what may be an unevaluated structure based upon personal experience rather than upon adequately investigated facts?

In its second role the Rorschach is becoming more extensively used as an experimental instrument, being applied to a wider range of problems. Here it is not the subjective impressions obtained by the clinicians which have been utilized so much as the scores which are recorded for the various Rorschach location, determinant, and content variables. It is in this aspect of the utilization of the Rorschach that we are most interested at the present time.

Since the research applications of the Rorschach deal essentially with scores, then we must of necessity be concerned with how these scores are obtained. The present study demonstrates that the examiner introduces a significant variance into the obtained scores which, as one illustration of the possible distortion, might tend to make nosological groups converge in their mean final Rorschach scores on various determinants. It may be assumed that experimental results containing variance introduced by the examiner would affect the theoretical constructs based upon the evidence. For example, studies based upon the Rorschach data collected by a single examiner would be contaminated by the particular personality characteristics of that examiner. It has been shown that examiners differ among themselves in regard to their effect upon the subjects. Studies based upon the Rorschach data collected by several examiners might be suspect for the same reason. In addition there is the further complicating factor of the differential reactions of individuals to the examiners. All these must be controlled before the Rorschach data can be most properly utilized in research.

The present study has demonstrated that examiner influence occurs on the Rorschach during the inquiry, and that its effect on the free association is minimal. It is therefore felt that attention should be given to the development of techniques designed to reduce the possibility of examiner influence. One such technique might be to give the subject the usual instructions prior to the administration of the test, but in addition to ask at that time that he give also the location and determinants. For example, he could be given the following instructions: "Tell me where you see it, and what suggests it to you." It is felt that the merits of this technique warrant consideration and experimental exploration.

It is not the purpose of this discussion to belittle the importance of the inquiry insofar as the results are utilized by the clinician in his handling of the problems presented by the individual patient. The verbalizations and content produced are often clinically rich in that they offer clues to the underlying dynamic processes with which the clinician is concerned. Rather, it is the fact that the complexities of the interaction of the subject and the examiner during the inquiry result in scores which are not only properties of the personality of the subject but also those of the examiner. Beyond the implications of this being a contaminatory factor in research applications, the findings have additional implications for the clinical interpretation of the Rorschach. The inquiry may be viewed not only as an interaction process between the subject and the examiner, but as a situation in which a second distinct perceptual process is experienced by the subject during the re-presentation of the Rorschach plates. This might be considered in itself as an additional unique perceptual experience. Systematic studies of the changes which occur on this re-presentation would be valuable since the changes from free association to inquiry within a given record might have considerable clinical significance. The findings of the present study offer some clues as to this process of change.

It was found that those individuals who changed the least from free association to final response following inquiry had a significantly higher amount of *M* and *FC* and lower number of *F* responses in their free-association records than did those who changed the most. This held true regardless of whether or not the examiner influence was controlled. The higher *M* suggests that those individuals with the least changes are possibly more comfortable in the interpersonal situation, relating more adequately to the examiner. They are less ready to interpret the situation as one of threat and consequently change relatively little in the shading determinants. Shading on the inquiry in the previous study [4] was shown to be related to the overt hostility of the examiner. The higher *FC* may be presumed to indicate that they relate better to external environmental situations and that their emotional reactions are more stable. The higher *M* and *FC* scores suggest that those individuals who in the Rorschach test situation change the least from free association to inquiry may be characterized as more rich and mature personalities. They possess a comparatively higher degree of psychic energy leading to a more complete and fuller initial response to the Rorschach blots. This is again suggested by the findings in regard to *F*. The persons changing the most had a significantly higher number of *F* responses on free association than did those who changed the least, being more constricted and showing greater affective suppression.

Finally, there is another crucial aspect of the relationship of the free association to the inquiry which has not been investigated. The question now arises as to their relative validity as measures of the personality characteristics of the individual. Which gives the more accurate picture of the person? Additional studies are planned to investigate this area.

Summary

1. Three areas of the relationship of the Rorschach free association and inquiry were investigated: (*a*) their similarity, (*b*) the comparative effect of the stimulus value of the examiner on each, and (*c*) the relationship of the personality characteristics of the subject to changes from free association to inquiry.

2. The subjects were 240 patients at the Detroit Mental Hygiene Clinic tested by 12 examiners, each testing 20 patients.

3. It was concluded that (*a*) the free-association record significantly differs from that of the inquiry, (*b*) the free association is relatively free of examiner influence and virtually all influences which occur are attributable to the inquiry, (*c*) the personality characteristics of those individuals changing the most differ significantly from those changing the least.

4. The implications of the findings for both the clinical and research utilization of the Rorschach were discussed. Of particular importance was the need for re-evaluation of past studies in light of the differential examiner influences on the inquiry. The need for studies of the relative validity of the free association and inquiry was stressed.

Received January 12, 1953.

References

1. Beck, S. J. *Rorschach's test.* Vol. I. *Basic processes.* New York: Grune & Stratton, 1944.
2. Cleveland, S. E. The relationship between examiner anxiety and the subject's Rorschach scores. Unpublished doctor's dissertation, Univer. of Michigan, 1950.
3. Gibby, R. G. The stability of certain Rorschach variables under conditions of experimentally induced sets: 1. The intellectual variables. *J. proj. Tech.,* 1951, **15,** 3–25.
4. Gibby, R. G. Examiner influence on the Rorschach inquiry. *J. consult. Psychol.,* 1952, **16,** 449–455.
5. Gibby, R. G., Miller, D. R., & Walker, E. L. Examiner influence on the Rorschach records of neuropsychiatric patients. *Amer. Psychologist,* 1952, **7,** 337-338. (Abstract)
6. Hutt, M. L., Gibby, R. G., Milton, E. O., & Pottharst, K. The effect of varied experimental "sets" upon Rorschach test performance. *J. proj. Tech.,* 1950, **14,** 181–186.
7. Milton, E. O. The stability of certain Rorschach variables under conditions of experimentally induced sets: 2. The stability of the human movement variable. Unpublished doctor's dissertation, Univer. of Michigan, 1949.
8. Pottharst, K. The stability of certain Rorschach variables under conditions of experimentally induced sets: 3. The effects of a complex set. Unpublished doctor's dissertation, Univer. of Michigan, 1950.
9. Sanders, R. The relationship between examiner hostility and subjects' Rorschach scores. Unpublished doctor's dissertation, Univer. of Michigan, 1950.

Journal of Consulting Psychology
Vol. 17, No. 5, 1953

The Use of the Rorschach in Involutional Melancholia

Frank J. Kobler
Loyola University

and Agnes Stiel[1]
Guidance Department, Catholic Charities, Chicago

The present investigation has for its purpose the analysis of the Rorschach records of 45 cases of involutional melancholia. Relatively little has been published about the Rorschach patterns for this type of mental abnormality. What is available is about depressive states generally, without much attempt to differentiate a specific pattern for involutionals. The evidence of the present study indicates that the typical involutional in this group is quite distinct from the individual presented in the Rorschach literature on depressed states under which involutional melancholia has been hitherto rather indiscriminately subsumed. These data then should represent an aid in a more adequate differential diagnosis.

The involutional psychosis is characterized by depressions occurring in middle and later life without evidence of any organically determined intellectual defects. The symptoms include agitation, uneasiness, unreality feelings, nihilistic delusions, anxiety, depression, insomnia, self-condemnation, and hypochondriacal elements. There is no history of previous attacks [5]. There are two types of involutional melancholia, the paranoid type in which there are delusions of persecution without a history of paranoid behavior, and the depressive type with typical signs of profound melancholia. While genetic and endocrine factors, particularly the latter, are important etiological factors, the social factors must not be neglected. Among the latter the factors of status and personal adequacy are of particular importance [3]. In involutional melancholia the self-concept is thought to undergo radical changes as a result of increasing age, menopausal changes, appearance changes, changes in energy and ambition, family disintegration, sickness, especially in relation to death, and a concern with the next life. For these people, the future is said to be essentially threatening and negative rather than stimulating and challenging.

The Rorschach data collated from previous studies leave something to be desired. As previously indicated most of the studies, specifically those of Rorschach [10], Guirdham [4], Varvel [11], Beck [1], Bochner and Halpern [2], Klopfer and Kelley [8], give factual information with regard to the general category of depressives but not adequately for involutionals. Rapaport [9], Johnson and Sherman [6], and particularly Kendig and Vernier [7] deal specifically with the involutional's response to the Rorschach. Of these studies that by Kendig and Vernier is easily the most systematic, carefully designed, and diagnostically useful that has appeared to date. In general there is little system in the way the facts are reported, in what is reported, and how the findings are worded. Almost all the investigators present their findings in descriptive and relative terms. Their data are virtually unamenable to exact statistical evaluation, much less cross-validation. An attempt will be made however to summarize briefly the facts that are available from these studies.

Rorschach [10] himself does not give the number of cases, Guirdham [4] has 161 depressives and 17 agitated depressives, Varvel [11] has 34 depressives and 14 agitated depressives, Beck [1], Bochner and Halpern [2], Klopfer and Kelley [8], and Johnson and Sherman [6] do not give the number of cases upon which their generalizations are based. Rapaport [9] used 7 cases, and Kendig and Vernier [7] so far have provided data on 25 depressives and 25 paranoid involutionals.

---

[1] The support of Ralph D. Bergen, Catholic Charities, Chicago, and Vincent V. Herr, S. J., Loyola University, and the statistical assistance of Robert G. Milligan, Loyola University, are gratefully acknowledged.

With regard to the number of responses, three of these studies [2, 8, 11] give no results, the remainder [1, 4, 6, 10] with one possible exception [9] indicate that the Rorschach responsiveness is diminished. Reaction time of these subjects is unknown from three studies [4, 9, 11], whereas the remainder indicate that it is prolonged without being much more specific. Bochner and Halpern [2] and Klopfer and Kelley [8] indicate that it will generally be sixty seconds or longer. In general one can say that all the studies are in substantial agreement that the *F%* is considerably increased over what is normally expected. In several instances *F%* is over 80. The number of *M* responses again finds rather general agreement among the investigators in that there is a reduced production in contrast with the normal. *M* is either absent or very low in patients of this type, although some exceptions are found [6, 8, 10]. All the studies likewise generally agree that the *C* response on the Rorschach is markedly reduced or absent. Such phrases as "few or none," and "essential absence of *C*," characterize the findings. Two studies [2, 8] speak of an increased *C′* in place of the more usual *C* scores. The four studies that indicate any findings about light and shading responses show that they are usually few in relation to what might normally be expected [1, 2, 4, 11]. The *W* factor in the test is generally decreased for this type of patient [1, 2, 6, 8, 11], although at least four studies suggest no decrease or an actual increase of *W* [4, 8, 9, 10]. This increase in *W* may however be for the neurotic type of depression rather than for the psychotic type. The findings with respect to the relations between *H* and *Hd* are unanimous, where present, in finding an increase of *Hd* over *H* [1, 4, 8, 10, 11]. It is also generally agreed by these investigators that the *A%* is increased [1, 2, 4, 8, 10, 11]. Two studies find a decrease in *A%*, but these two are specifically for involutional patients [6, 9]. Three studies report on the original response [4, 8, 10]. In all instances they are low or absent, and if present they tend to be bizarre [6]. The findings with regard to the popular responses are inconsistent. Six of the eight studies report such data, of which two [4, 8] suggest an increase and three [6, 9, 11] suggest a decrease.

The preliminary findings of the Kendig and Vernier study warrant special comment. They found that the typical depressive pattern was absent in their cases. Instead their research showed an organic pattern, but was distinguished from the usual organic profile by the adequate form level, adequate populars, and an average total number of responses above fifteen. They likewise report a prolonged reaction time. Their cases show an increased *W* and *F%*, a decreased *M*, *D*, and *O* responsiveness, and few shading responses.

In a relatively recent study by Young [12] 13 cases of involutional melancholia were given the Rorschach. He finds a distinct pattern for these cases differing from other types of mental abnormality. Response average was low, *W* was increased, *M* reduced or absent, *C* was absent or variable, and animal and anatomical responses predominated. Qualitative factors were stressed. The more important ones included: depreciatory remarks, stereotypy, paucity and rigidity of thought, and irrelevant responses.

It seems clear from a brief review of the literature that most of the evidence that is available on the quantitative structuring of the Rorschach in this area has to do with the records of depressives, both neurotic and psychotic. Since one of the basic elements in the involutional picture is that of depression, it has generally been assumed that the depressive diagnostic pattern revealed in the literature could be used to make a correct diagnosis of involutional psychosis. If the depressive element was the key element underlying these several mental illnesses presumably the Rorschach responses would be affected in a substantially similar way.

Subjects

The subjects used in this study were 45 carefully diagnosed female involutional psychotics admitted to a state hospital. Twenty-two of the cases were diagnosed as involutional psychosis proper, and 23 cases were diagnosed as involutional psychosis, paranoid type. The mean age of the subjects was 49.3 years with a standard deviation of 4.4. The mean number of years of education was 8.4 with a standard deviation of 4.8 years. The mean intelligence quotient for 28 of these patients, taken from the Wechsler-Bellevue, was 92.3. There were no statistically significant differences between the mean age and the mean years of education of the two diagnostic subgroups. Thirty-five of the subjects were housewives and 37 of them were married. All the patients were oriented in the usual psychiatric sense. Thirty-four of the patients had accelerated motor behavior, 9 were normal in this respect, and 3 were retarded.

Discussion of Results

The present data indicate that both the quantitative and the qualitative picture provided by these involutional cases is not even remotely similar to the usual depressive picture, whether neurotic or psychotic. Setting up involutional psychosis as a separate disease entity, as is argued by most psychiatric authorities, finds considerable support in the data of the present involutional group. An examination of Table 1 presents some of the facts that can be used in support of such a contention. Certain of these findings are worth emphasizing.

The involutionals certainly do not produce the number of achromatic responses in proportion to the chromatic ones which are generally to be expected in a depressed patient. For the

Table 1

Rorschach Test Characteristics of a Group of Forty-five Cases of Involutional Melancholia with Contrasted Subdiagnostic Groupings of Paranoid and Depressed Cases

Score	Paranoid ($N=23$) Mean	Paranoid ($N=23$) *SD*	Depressed ($N=22$) Mean	Depressed ($N=22$) *SD*	Total ($N=45$) Mean	Total ($N=45$) *SD*	Total ($N=45$) Med.	Total ($N=45$) *Md.D*	χ^2 *
Total *R*	10.89	3.56	9.23	3.73	10.08	3.74	10.20	3.02	.01
Total time	426.0	173.6	442.0	152.7	434.0	163.8	412.0	135.2	.02
Time per *R*	39.30	17.70	50.40	22.70	44.70	21.00	40.20	16.31	.20
Av. reac. time Achro. cards	16.35	9.80	16.55	10.45	16.45	10.10	13.60	7.51	.22
Av. reac. time Chro. cards	13.75	7.90	15.65	9.15	14.80	8.60	13.20	6.52	1.07
W	5.39	2.08	5.32	2.28	5.36	2.18	5.00	4.89	.21
D + *d*	5.39	2.90	4.32	2.93	4.87	2.96	4.00	9.53	.23
Dd + *S*	0.17	0.37	0.27	0.67	0.22	0.56	0.00	0.22	.10
F	7.22	3.11	5.59	2.66	6.42	3.01	6.33	2.39	2.53
F%	69.30	21.20	60.90	22.10	65.30	22.10	64.50	16.20	1.81
M	0.17	0.37	0.14	0.35	0.16	0.36	0.00	0.16	.01
FM	0.30	0.55	0.64	0.37	0.47	0.93	0.00	0.47	.58
m	0.17	0.48	0.05	0.22	0.11	0.39	0.00	0.11	.23
FC	0.39	0.92	0.23	0.42	0.31	0.72	0.00	0.31	.08
CF	1.00	1.24	1.17	1.56	1.09	1.43	0.00	1.09	.02
C	0.09	0.28	0.14	0.35	0.11	0.32	0.00	0.11	.002
Sum *C*	1.33	1.35	1.53	1.76	1.42	1.57	1.00	2.65	.01
Sum *c*	0.52	0.65	0.64	0.78	0.58	0.71	0.00	0.84	.02
Sum *C'*	0.56	0.72	0.86	1.17	0.71	0.98	0.00	0.71	.02
Sum *K*	0.52	2.04	0.32	0.76	0.42	1.54	0.00	2.53	.01
A%	42.30	22.45	61.80	21.98	51.80	24.27	52.70	18.06	2.84
P%	2.04	1.12	2.73	1.17	2.38	1.20	2.46	0.75	.20
O%	0.70	1.08	0.05	0.22	0.33	0.87	0.00	0.38	5.92
Cards 8, 9, 10/*R*	29.90	6.00	31.41	7.96	30.82	7.00	30.78	5.85	1.79
M:C equals 0:0 to 0:2	70%		64%		67%				.18
FMm:FcC' 0:0 to 0:2	65%		55%		60%				.53
$H + A > Hd + Ad$	74%		86%		80%				.30
$W(3x) \rightarrow M$	91%		86%		89%				.003

* With 1 *df*, χ^2 must reach 3.84 to be significant at .05 level.

present group the sum of the chromatic responses was actually larger than the sum of the achromatic responses. The emotional responsiveness of the involutionals is indicated further in the predominance of the *CF* and *C* over the *FC* responses. There were no *M* responses in the group of any consequence. Of the total of 45 cases only 6 had any *M* responses, and each of these had only one *M*. The movement response is obviously not produced by the involutionals. It may be as Rorschach suggests that people with accelerated motility produce few *M* and more color responses. These three factors then characterize this group. The average *F%* is high in this group but it is also considerably lower than has been reported, without data, in much of the literature. The responses that generally accounted for the lowered *F%* were the *Fc*, *c*, and *CF* responses. Of additional considerable importance in support of the rejection of the usual depressive profile for the involutionals is the almost total absence of the *F*+ response. Form responses of involutionals are characteristically mediocre or poor. The proportion of *W* responses is approximately the same as for other types of "depressed" patients, perhaps slightly lower. Response and reaction times of the involutionals however tend to be within the normal range rather than exaggerated as in depressives. The involutionals characteristically give quick responses even though there is considerable clinical anguish and despair. On the other hand it often took literally hours to get these patients to continue responding between certain cards. It is also possible to say in the light of the evidence that these patients are not typically stereotyped as has been reported for other types of depressives. This can be determined from the type and from the variety of the content.

The findings with respect to the proportion of the *H*+*A* responses to the *Hd*+*Ad* reponses was clear-cut. In almost 100 per cent of the cases the sum of the *H*+*A* responses was equal to or exceeded one-half the sum of the *Hd*+*Ad* responses. It seems clear that as far as the involutional records go, the human and animal part responses are not excessively proportionate to the human and animal whole reponses. Popular responses were also considerably lower than expected for the depressives. The average *P* was slightly over two rather than the expected five. However the number of additional populars was high. Here the patient gave the "popular" response without the complete detail required for popular scoring. This neglect in "completing" properly the popular response, when given, and the unusual verbalizations often associated with the populars, was characteristic of the individual *P* response.

Despite readily observable intense clinical anxiety in the majority of these patients, the appearance of *K*, *k*, and *m* responses was extremely rare. The mean frequency of all such responses was less than one-half a response. The anxiety shown in these protocols manifests itself in ways different from the anxiety signs of the typical psychoneurotic [4]. The differences in the quantitative responses to the chromatic and achromatic cards are likewise not characteristic of the depressive picture. For the present data the increased reaction time to achromatic cards over the chromatic cards is not a typical indicator of depression. The popular responses to chromatic cards do not exceed the popular responses to achromatic cards. The number of rejections to achromatic cards does not exceed the number of responses to the achromatic ones.

It is worth noting that, statistically, there are no differences of any consequence between the paranoid and the depressive subgroups of the total involutional sample. This is true for all the test factors susceptible to such quantitative analysis. Using chi square, and applying the Yates correction in all instances, as the basis for the analysis, none of the differences approach an acceptable probability level. The *O%* responses with a chi square of 5.92 reaches a p of .01 with one degree of freedom. It is reasonably questionable whether this one statistically significant difference, among so many others, has any importance. For all practical purposes, and from the quantitative point of view, the subgroups do not differ, and the total group is homogeneous. In addition since the use of a mean and the standard deviation might, in relation to the assumptions underlying their use, give questionable results, the medians and median deviations were calculated for all the data. No differences of any consequence for these two different measures of central tendency and dispersion were found for these data.

This is clear from the last column of Table 1 where the comparable statistics are given.

Qualitative Analysis

An analysis was made of the 45 involutional protocols to determine the nature and the extent of the qualitative factors that characterized the group. It was thought that these might be even more useful than the more customary quantitative factors used in arriving at a differential diagnosis. The approach was inductive and empirical. The records were read carefully by the investigators and fourteen factors were isolated. Eight of these factors had usefully large loadings.

The use of words and phrases indicating uncertainty and the need for reassurance characterized the entire group. Such expressions as "is it right or wrong," "maybe I said it wrong," "I want to say the right thing," all illustrate a process of thought and feeling that reappeared regularly and consistently. Among the paranoid subgroup such expressions appeared in the records of 61 per cent of that group, with an average of 4.3 repetitions in the records where it appeared. Fifty per cent of the depressives had an average of 2.2 such phrases in those protocols. Thus over one-half of the total group of involutionals used such expressions on the average of three and one-half times per record.

The total number of rejections of the cards represents an important aspect of the records qualitatively. Sixty-one per cent of the paranoids had a total of 32 rejections for an average of 2.3 rejections per person doing any rejecting. Fifty-nine per cent of the depressives had a total of 46 rejections, or an average of 3.5 rejections for each patient in this group. Rejection rates are thus approximately the same for both subgroups and are found in 60 per cent of the cases of involutionals. The most frequently rejected cards are six, seven, and nine. The least frequently rejected cards are one, five, and eight. All cards showed some rejections. The order of frequency of rejections was: cards six, nine, seven, ten, three, four, two, eight, one, and five.

Considerable self-reference characterizes the group. Several examples of this type of reference will help to explain what is meant by this term. The patients said such things as "am I crazy," "can't my husband take me out," "I don't want to die," "you are leading me to the gallows." There were numerous hypochondriacal complaints. Of a total of 23 paranoids 13 or 57 per cent used such forms of self-reference. It was used 2.6 times on the average for this group. The depressive subgroup of 22 patients showed 45 per cent using such self-reference for an average of 3.1 times. One-half of the total group of involutionals used self-reference a total of 65 times for an average use of 2.8 times for the users.

Sexually toned responses were frequently used, especially among the paranoid patients. Typically the sex response was evasive and general. Examples are: "newborn," "below a woman's stomach," or "this one makes me hot." Some responses were more specifically sexual, such as "breast" and "vagina." Such responses occur more than twice as frequently among the paranoids than they do among the depressive involutionals. Sex responses occurred in 57 per cent of the paranoid protocols for an average of 3.9 appearances. Only 23 per cent of the depressives gave such responses for an average of 2.4 times in the protocols where they were found.

Perseveration was markedly apparent in the protocols of these involutionals. Eleven out of 23 paranoids, or 48 per cent, showed perseveration 42 times, or 3.8 times on the average for the persons represented. The involutional depressed group showed 41 per cent perseverating for an average of 2.7 times per protocol. Thus approximately one-half of the total group showed perseveration on the average of 3.3 times in these records.

There are two types of perseveration that may logically be distinguished in these records. One type of perseveration is the usual kind which appears in the presence of the F− response. The other type usually appears three or more times successively but is not accompanied by the F−. This latter type is decidedly more frequent and typical here. Of a total of 20 out of 45 cases that showed any perseveration, the latter type was found in 70 per cent of the cases. Such things as body, flower, Xray, bones, animals, just colors, are typically perseverated.

The characteristic perseveration seen here reflects a sort of mental or associational "stickiness" without impairment of intellectual control. This may be a temporary limitation on perceptual variability, or an inability, because of intense anxiety, to depart easily from the quickly seen or accepted. It is a perseveration of fear, of avoidance, of vacuity, rather than one of dissociation or deterioration.

Many of the women had the practice of ascribing undesirable characteristics to the blots. They used such words as, "mean, evil, frightful, having horns, like the devil, unclean, and threatening." In the paranoid group such expressions occurred 18 times in 5 protocols, or in 22 per cent of the cases. In this group the average occurrence was 3.6 times. The depressives showed this in twice as many, or in 45 per cent of the protocols, the average occurrence in these records being 1.6 times. For the total group this type of response occurred in one-third of the cases with an average frequency of 2.3 times.

F— responses were infrequently found in these involutional protocols. Only ten out of a total of 45 protocols showed such a response. Of this small group 90 per cent are found among the paranoids. The records that have this type of response show it on the average of 3.2 times per protocol.

Paranoid references are seen in one-half of the paranoid group and not at all in the depressive group. This is one qualitative feature that definitely distinguishes the two groups, as might be expected. The 52 per cent of the paranoid subgroup who made such responses gave them on the average of 2.3 times per protocol.

Certain facts stand out in evaluating the qualitative response characteristics of this involutional group. In the first place the subclassification of the group into paranoids and depressives reveals certain important differences between them, as well as important similarities. Presumably these differences should have some value for differential diagnosis within the group. The subgroups are homogeneous with respect to expressions of uncertainty, need for reassurance, self-reference, perseveration, and rejections. Depressives ascribe undesirable characteristics more frequently, while sex, paranoid, and *F*— responses occur more often among the paranoid involutionals.

Summary

The protocols of 45 involutional melancholics were analyzed quantitatively and qualitatively. The findings of this analysis were contrasted with those of similar studies to determine differences and reliable diagnostic indications. The results indicate a pattern in involutionals consistently dissimilar from the usual depressive picture, whether neurotic or psychotic. Involutional patterns can be discerned with reasonable diagnostic accuracy when both quantitative and qualitative factors are employed. Quantitative factors differentiate the involutional from other types of mental disorders. Significance figures show that only qualitative factors aid in the differential diagnosis of paranoid and depressive involutionals.

Received July 16, 1953.
Early publication.

References

1. Beck, S. *Rorschach's test.* Vol. 2. *A variety of personality pictures.* New York: Grune & Stratton, 1944.
2. Bochner, Ruth, & Halpern, Florence. *Clinical application of the Rorschach test.* New York: Grune & Stratton, 1945.
3. Cameron, N. The functional psychoses. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944.
4. Guirdham, A. Diagnosis of depressions by the Rorschach test. *Brit. J. med. Psychol.*, 1936, **16** 130-148.
5. Henderson, D., & Gillespie, R. D. *Textbook of psychiatry.* (7th Ed.) London: Oxford Univer. Press, 1950.
6. Johnson, C., & Sherman, J. Clinical significance of the Rorschach test. *Amer. J. Psychiat.*, 1948, **104**, 730–737.
7. Kendig, Isabelle, & Vernier, Claire. Rorschach patterns in involutional melancholia. *Amer. Psychologist,* 1948, **3**, 282. (Abstract)
8. Klopfer, B., & Kelley, D. M. *The Rorschach technique,* Yonkers, N. Y.: World Book Co., 1946.
9. Rapaport, D. *Diagnostic psychological testing.* Chicago: Year Book Co., 1946. 2 vols.
10. Rorschach, H. *Psychodiagnostics.* New York: Grune & Stratton, 1942.
11. Varvel, W. The Rorschach test in psychotic and neurotic depressions. *Bull. Menninger Clin.*, 1941, **12**, 5-12.
12. Young, R. J. The Rorschach diagnosis and interpretation of involutional melancholia. *Amer. J. Psychiat.*, 1950, **106**, 748–749.

Journal of Consulting Psychology
Vol. 17, No. 5, 1953

A Cross-Validation of Suggested Rorschach Patterns Associated with Schizophrenia[1]

Harold Rubin and Murray Lonstein
Veterans Administration Hospital, Lebanon, Pennsylvania

In a recent study Thiesen [2] has identified several quantitative combinations of Rorschach structural variables as being associated with schizophrenia. Five tentative patterns are offered as being of diagnostic value in differentiating between a patient group diagnosed as schizophrenics and a "normal" control group.

Such a pattern-analytic approach to the Rorschach may succeed in isolating combinations of Rorschach variables which may identify certain pathological groups. This approach may be regarded as a step closer to establishing the validity of discrete Rorschach variables.

Thiesen [2] is aware of the limitations of his findings and points out the need for cross-validating the patterns which have emerged from his investigations. In the present study, the authors have attempted such a cross-validation by applying Thiesen's findings to another schizophrenic sample.

Procedure

The subjects used in the present study were taken from a group of hospitalized male patients diagnosed as schizophrenic by the psychiatric staff of a Veterans Administration Hospital. Rorschach tests were administered individually to the patients by the authors who utilized Beck's revised scoring system [1]. Only those records having ten or more scorable responses were retained for this study. A total of 42 protocols was obtained.

The Rorschach variables which comprised the patterns suggested by Thiesen [2] as being associated with schizophrenia were selected from the protocols. The numerical value of each variable was assigned to one of three categories of magnitude, "low, medium, and high," as established by Thiesen on a normal control group. Tables were then constructed which described the frequency of occurrence of the suggested patterns in the present sample. These data were analyzed by the chi-square technique.

Results

In Table 1 are listed the five patterns suggested by Thiesen. Here comparisons are made between the findings of the present study and Thiesen's control group. In patterns *A, B, C,* and *E* no significant differences were obtained, while pattern *D* was significant at the .05 level.

Table 1

Patterns Associated with Schizophrenia

Pattern	Description	Frequency		Chi square
		Controls (N = 157)	Patients (N = 42)	
A	*An*: High *Sex*: High	1 (0.6%)	1 (2.3%)	.02
B	*F*+%: Low *Z* Score: Low	2 (1.3%)	2 (4.7%)	.66
C	*FC*+: Low *M*: Low *A*%: Low	0 (0.0%)	1 (2.3%)	.50
D	*F*+%: Low *P*: Low *A*%: Low	1 (0.6%)	3 (7.1%)	4.20
E	*DW*: High *FC*+: Low	1 (0.6%)	1 (2.3%)	.02

Table 2 indicates the frequency of occurrence of one or more of the patterns in both the present sample and Thiesen's control group. It is observed that 7 of the 42 patients or 16.7% were characterized by one or more patterns, whereas 5 of the 152 controls or 3.2% had one or more of these patterns.

---

[1] From the Veterans Administration Hospital, Lebanon, Pennsylvania.

There was only a single occurrence of more than one pattern in any record, a combination of patterns *A* and *C*.

Table 2

Frequency of Pattern Occurrence in Control and Patient Populations

Group	Number of Patterns None	 One or More	Total
Controls	152 (96.8%)	5 (3.2%)	157
Patients	35 (83.3%)	7 (16.7%)	42

Chi square = 10.63 ; p < .01.

Table 3 compares the frequency of occurrence of one or more of the patterns between Thiesen's schizophrenic group and the present schizophrenic group. The authors' findings indicate 7 of the 42 records or 16.7% have one or more patterns, while Thiesen reports that 29 of his 60 patients or 48.3% fall in this category. This difference yielded a chi-square value of 9.50, which is significant at the .01 level of confidence.

Table 3

Frequency of Pattern Occurrence in Both Patient Populations

Group	Number of Patterns None	 One or More	Total
Patients (Authors')	35 (83.3%)	7 (16.7%)	42
Patients (Thiesen)	31 (51.7%)	29 (48.3%)	60

Chi square = 9.50 ; p < .01.

Discussion

In contrast to Thiesen's results, which showed significant differences of all five patterns between his patient group and the control group, the authors found that only pattern *D* appeared to discriminate between their patient group and the controls. Because of the small frequency in each cell, this finding may be of doubtful significance.

Thiesen reports the presence of one or more patterns in 48.3% of his schizophrenics as compared with 3.2% of his controls. This would be a valuable diagnostic finding. In this cross-validation, however, only 16.7% of the schizophrenics were found to have one or more patterns. This shrinkage lessens the clinical utility of these patterns.

Table 4

Frequency of Occurrence in Magnitude for Variables Comprising Patterns *A*, *B*, *C*, *D*, and *E* in the Present Schizophrenic Sample ($N = 42$)

Variable	Low	Medium	High
An	19	12	11*
Sex	36	3	3*
F+	17*	11	14
Z	10*	31	1
FC+	31*	5	6
M	16*	24	2
A%	8*	18	16
P	22*	18	2
DW	37	0	5*

* Magnitudes of the variables specified in the patterns.

Examination of the data in Table 4 shows that of the individual Rorschach variables comprising patterns *A*, *B*, *C*, and *E* (those patterns not found to be discriminatory) only three variables would appear by inspection to characterize the present group of schizophrenics. These are low *FC*+, low *M*, and low *P*.

Although the present schizophrenic group differed significantly from the control group (Table 2), it also differed significantly from Thiesen's schizophrenic group (Table 3). This discrepancy indicates a lack of homogeneity in the total schizophrenic population sampled in both studies and underlines the limited value of the pattern approach.

Summary

A cross-validating study was carried out on Thiesen's suggested Rorschach patterns associated with schizophrenia. Forty-two schizophrenic Rorschach records were analyzed, and these patterns failed to characterize this group.

Received January 20, 1953.

References

1. Beck, S. J. *Rorschach's test.* Vol. I. *Basic processes.* New York: Grune & Stratton, 1949.
2. Thiesen, J. W. A pattern analysis of structural characteristics of the Rorschach test in schizophrenia. *J. consult. Psychol.*, 1952, **16**, 365-370.

Journal of Consulting Psychology
Vol. 17, No. 5, 1953

Differences Between the TAT Responses of Negro and White Boys

Paul H. Mussen
The Ohio State University

The extent to which an individual's projections may be determined by group membership and sociocultural factors as well as by personal idiosyncratic factors has had little systematic study. It seems reasonable to assume that there will be "consistent differences between the fantasy productions of individuals who belong to, or have been socialized in, different social groups. Thus, a certain amount of the variation in any TAT production can be accounted for by the fact that the individual has grown up in a given milieu or social role" [5, p. 16]. Henry [4] found that there were systematic differences between the fantasy productions (TAT stories) of Hopi and Navaho children which reflected important differences between their cultures.

In their study designed to test the assumptions underlying the Thompson TAT, Riess, Schwartz, and Cottingham showed that "Negroes and whites in the North produce stories that differ insignificantly in length regardless of whether the stimulus material is Negro or not and regardless of the color of the experimenter" [7, p. 708]. Using a reliable measure of the number of ideas contained in the stories, these same investigators also concluded that these two groups "produce ideas in the TAT situation that are not statistically different in number regardless of whether the stimulus material is Negro or white and regardless of the color of the experimenter" [8, p. 396].

Although these findings make it clear that differences between Negroes and whites in the quantity of production in response to the TAT are insignificant, they contribute nothing concerning the possible dynamic differences between the two groups' fantasy productions. Since Negroes and whites in America grow up in different social milieux and consequently are subject to different cultural pressures, it may be predicted that their responses to projective materials will differ. The present study was designed to investigate differences between the contents of the TAT stories of Negro and white children in New York City.

Method

In connection with a study of personality and social factors related to changes in children's attitudes toward Negroes, the author [6] administered 13 TAT cards to lower-class Negro and white boys from New York City who were vacationing at a social agency's summer camp.[1] The cards used were: 1, 2, 3, 4, 6, 7, 8, 12, 13B, 14, 16, and 18GF of the Murray TAT and another card, not part of the regular series, which depicts a smiling woman leaning over a baby's crib and giving the baby a pacifier, while a boy about eight years old stands behind the crib.

The data of the present study were derived from the protocols of 50 Negro and 50 white boys of at least normal intelligence. The subjects were matched in age so that there were equal numbers of white and Negro boys at each age level between nine and fourteen.

All TAT stories were analyzed in accordance with a scheme suggested by Stein [9] and modified somewhat by the author. The scheme involves counting the needs and press of the heroes of the stories, the assumption being that the storyteller has identified with the hero; the hero's needs are the same as the child's needs; the press that impinge upon the hero are the same as the press that affect the child telling the story.

Twenty-eight needs and 22 press categories were employed in the analysis of the protocols. A score for each subject for each personality variable (i.e., need and press) was derived by adding the number

---

[1] The author wishes to acknowledge his indebtedness to the Community Service Society of New York for granting permission to carry out this study at their boys' camp.

Table 1
Number of White and Negro Subjects above the Mean on 50 TAT Variables

TAT Variable	Definition of Variable	High Whites N	High Negroes N	Chi-Square Value	p
*n Achievement	To work at something creditable	24	6	15.40	<.01
n Acquisition 1	To work for money, possessions	26	19	1.98	.10—.20
n Acquisition 2	To steal, cheat, swindle	26	22	.64	.30—.50
n Affiliation 1	Enjoying parent's company	29	24	.99	.30—.50
*n Affiliation 2	To establish, maintain friendly relations	19	9	4.95	.02—.05
n Affiliation 3	To fall in love, marry	24	15	3.41	.05—.10
*n Aggression 1	Self-defensive aggression	22	6	12.66	<.01
n Aggression 2	Physical, asocial aggression	29	27	.16	.50—.70
*n Aggression 3	Emotional, verbal aggression	15	26	4.98	.02—.05
n Aggression 4	Destruction, smashing, breaking	12	8	1.00	.30—.50
*n Aggression (kill)	Killing in anger	30	14	10.36	<.01
n Autonomy 1	To escape, avoid restraint	13	20	2.21	.10—.20
n Autonomy 2	To resist coercion	31	33	.17	.50—.70
n Autonomy 3	To misbehave, be unruly	25	35	3.17	.05—.10
n Deference 1	To submit unwillingly to another	13	17	.76	.30—.50
n Deference 2	To comply willingly, obey	29	27	.16	.50—.70
*n Deference 3	To express admiration, respect	14	3	8.58	<.01
n Dominance	To lead, manage, influence others	23	18	1.03	.30—.50
n Exposition	To inform, explain	26	24	.16	.50—.70
n Nurturance 1	To be kind, helpful to parents	13	18	1.17	.20—.30
*n Nurturance 2	To be kind, helpful to others	29	18	4.84	.02—.05
n Nurturance	To seek food, drink	17	19	.17	.50—.70
n Passivity	To enjoy relaxation	11	20	3.78	.05—.10
n Playmirth	To indulge in pleasurable activity	22	26	.64	.30—.50
n Recognition	To seek approval, prestige	16	9	2.60	.10—.20
n Regression	To think of early childhood	6	7	.09	.70—.80
n Succorance	To seek aid, sympathy	19	22	.37	.50—.70
*n Understanding	To think, reflect, speculate	18	32	7.84	<.01
p Acquisition	Hero robbed, swindled	20	21	.04	>.90
p Affiliation 1	Hero and parents united	23	19	.65	.30—.50
*p Affiliation 2	Friendly relations established with hero	30	16	7.88	<.01
p Affiliation 3	Love for hero expressed	21	13	2.85	.05—.10
p Aggression 1	Retaliation against hero	10	15	1.34	.20—.30
p Aggression 2	Hero punished	23	18	1.03	.30—.50
*p Aggression 3	Hero assaulted, injured	10	28	13.70	<.01
*p Aggression 4	Hero hated, quarreled with	15	31	10.30	<.01
*p Deference 1	Hero's leadership followed	18	7	6.46	.01—.02
*p Deference 2	Hero admired, respected	34	16	12.92	<.01
p Dominance 1	Hero forced to do something	24	31	1.98	.10—.20
p Dominance 2	Hero prevented from doing something	17	26	3.30	.05—.10
p Dominance 3	Hero persuaded, pleaded with	24	31	1.98	.10—.20
p Example	Hero influenced	11	17	1.78	.10—.20
p Lack	Hero lacks something he needs	16	14	.19	.50—.70
p Loss	Hero loses something	29	23	1.34	.20—.30
p Loss (parents)	Hero's parents die	28	24	.64	.30—.50
p Nurturance 1	Parents help hero	20	28	2.58	.10—.20
p Nurturance 2	Hero helped, encouraged	19	19	0.00	1.00
p Physical danger	Hero's well-being threatened	27	19	2.58	.10—.20
p Physical danger (parents)	Hero's parents endangered	23	20	.44	.50—.70
*p Rejection	Hero rejected, scorned	18	7	6.46	.01—.02

* Significant at 5% level or better.

of times it appeared in the thirteen stories. A list of the needs and press scores used, together with brief definitions of them, is found in Table 1. In order to eliminate bias, all scoring was done "blind," that is, independently of knowledge of the subject's race.

Results

The mean score for the entire group of 100 subjects for each of the 50 TAT scores was determined. Subjects having scores above the mean score for a given TAT variable were considered "high" in that particular need or press; subjects having scores below the mean score for a given TAT variable were considered "low" in that variable. Chi-square tests were used to determine whether or not there was a significant difference between the number of white and Negro subjects scoring high on each of the TAT variables.

Of the 50 chi squares computed, 14 yielded chi-square values which were significant at the 5% level or better. Table 1 lists the TAT variables, the number of Negroes and whites high in the variable, the chi-square value obtained, and the level of confidence. There is a considerably greater number of significant differences than could be accounted for on the basis of chance alone.

Discussion

Examination of Table 1 makes it clear that there are many significant and consistent differences between the contents of Negro and white boys' responses to the TAT.

The minority group understandably sees the general environment as more hostile than does the white majority group. Thus, a significantly greater number of Negro than white boys tell many stories in which the hero is hated, scolded, or reprimanded (*p* Aggression 4) or is the victim of physical assault (*p* Aggression 3). Despite this, however, their own aggressions seem to take a milder form than the white group's. Although they indulge in more verbal aggression and hostile thinking (*n* Aggression 3), they are relatively less concerned with extreme aggressive expression (*n* Aggression — kill) and self-defensive aggression (*n* Aggression 1) than are the white boys.

The Negro's view of the environment as essentially inhospitable, also observed by Cottingham [1], is further reflected in the finding that a relatively small number of these subjects see themselves as establishing and maintaining 'friendly relations (*n* Affiliation 2), respecting others (*n* Deference), or being kind and considerate of others (*n* Nurturance 2). In the same way, they infrequently see themselves as being respected (*p* Deference 2), followed, or obeyed by others (*p* Deference 1).

Their perception of the world as hostile and threatening may also imply feelings of inferiority and helplessness and an attitude of indifference. The broad effects of such an attitude may be seen not only in the Negro group's relatively infrequent use of extreme forms of aggression but also in the relative paucity of responses indicating striving for accomplishment and success (*n* Achievement), together with an emphasis on generally inactive pursuits such as thinking, reflecting, speculating (*n* Understanding).

By contrast, the white group considers the world to be predominantly friendly and they respond to it with more warmth and friendliness. Thus the heroes of this group's stories are leaders and respected individuals significantly more frequently than the heroes of the Negro boys' stories. Moreover, they are interested in establishing and maintaining friendly relations and are more often depicted as kind and considerate of others. Since they are relatively secure and unthreatened in the world at large, they feel free to strive to achieve something creditable, to verbalize their extremely hostile fantasies, and to defend themselves against aggression.

Feelings of rejection, generally by the mother, (*p* Rejection) are less frequent in the stories of the Negro than of the white boys. This may be due to the dominance of maternal authority and consequent feelings of identification with the mother in Negro families [2], or to more general feelings of security in the family situation resulting from the highly permissive treatment of very young children by lower-class Negro mothers [3]. Since it seems that they feel less rejected at home than the whites do, the Negro subjects' attitudes of indifference must be attributed to their perceptions of the world at large as generally hostile, rather than to feelings of insecurity in their earlier familial relationships.

Regardless of what their determinants are, feelings of rejection occur significantly more frequently in the protocols of the whites than of the Negroes. The relatively large number of extremely hostile attitudes found in this group's stories is very probably related to these feelings of rejection.

Evidence from the present study clearly supports the assumption that there are major differences in the fantasy productions of the two sociocultural groups studied. In addition to personal dynamic factors, the historical, economic, and social forces impinging on the group are reflected in the stories told by the group members. Apparently a certain amount of variation in these projective responses may be attributed to the individual's cultural background. Some aspects of his fantasy productions may be primarily determined — and most parsimoniously explained — by the individual's group membership and only secondarily by his unique history of interpersonal experiences.

It follows that interpretation of an individual's responses to projective materials must take account of his cultural milieu. A great number of responses of a certain kind may be quite unusual and perhaps indicative of maladjustment or unresolved conflicts and tensions if they are given by a member of one group, but usual or "normal" if given by a member of another group. The extent of usualness or deviation of projective responses must be evaluated from the point of view of the individual's group membership and cultural background.

Summary

The TAT protocols of 50 white and 50 Negro lower-class boys were analyzed and scored in terms of 50 categories of needs and press. There were significant differences between the numbers of members of the two groups with high scores (above the mean for all 100 subjects) in 14 needs and press.

The Negro subjects' stories contained significantly more incidents of aggressive press from the environment and mild, verbal aggressive expression by heroes. Compared with the whites, they showed less interest in establishing and maintaining friendly relations, or being kind to, or respecting others. The Negro boys' attitudes of indifference were further shown in their infrequent use of *n* Achievement and their emphasis on essentially inactive pursuits such as thinking and speculating.

The white boys seem to suffer more from feelings of rejection in the family and more frequently express extreme hostility in their fantasies. On the other hand, they see others as respecting them and following their leadership and they respond to the generally favorable social situation with establishing friendly relations, being considerate of others, and striving to achieve something creditable.

There are evidently major differences in the fantasy productions of individuals from different social groups. Dynamic interpretations of an individual's projective responses can be made meaningfully only if variations primarily attributable to his cultural background are taken into account.

Received February 5, 1953.

References

1. Cottingham, A. A study of the use of the Murray and Thompson modified TAT with a group of Negro and white subjects. Unpublished Educational Research Seminar report, School of Education, College of the City of New York, 1950.
2. Dai, B. Some problems of personality development among Negro children. In C. Kluckhohn, & H. A. Murray (Eds.), *Personality*. New York: Knopf, 1948. Pp. 437-458.
3. Davis, A., & Havighurst, R. J. Social class and color differences in child-rearing. *Amer. sociol. Rev.*, 1946, **11**, 698-710.
4. Henry, W. E. The Thematic Apperception Technique in the study of culture-personality relations. *Genet. Psychol. Monogr.*, 1947, **35**, 3–315.
5. Lindzey, G. Thematic Apperception Test: Interpretive assumptions and related empirical evidence. *Psychol. Bull.*, 1952, **49**, 1–25.
6. Mussen, P. H. Some personality and social factors related to changes in children's attitudes toward Negroes. *J. abnorm. soc. Psychol.*, 1950, **45**, 423–441.
7. Riess, B. F., Schwartz, E. K., & Cottingham, A. An experimental critique of assumptions underlying the Negro version of the TAT. *J. abnorm. soc. Psychol.*, 1950, **45**, 700–709.
8. Schwartz, E. K., Riess, B. F., & Cottingham, A. Further critical evaluation of the Negro version of the TAT. *J. proj. Tech.*, 1951, **15**, 394–400.
9. Stein, M. I. *The Thematic Apperception Test.* Cambridge: Addison-Wesley Press, 1948.

Journal of Consulting Psychology
Vol. 17, No. 5, 1953

The Use of Sounds in a Projective Test[1]

Harry A. Wilmer[2] and May Husni
Stanford University

An auditory apperception test has previously been reported by Stone [2] in a preliminary report. This test consists of 10 records of various human, animal, and mechanical sounds. The only other type of study in the literature of an auditory test is that of the Verbal Summator or Tautaphone described by Skinner [1].

Considering the importance of sound phenomena in human experience and the paucity of experimental studies, it was decided to devise a brief collection of sounds that could be used for clinical or experimental purposes. After a preliminary study [4, 5] in which 75 sounds were tested with 117 subjects, the present series of sounds was selected as being most productive or leading to the most interesting associations. A selection of 21 brief sound recordings has been played to subjects to study their response to sound perception. The need for a projective test using this sensory modality to supplement visual projective tests is apparent. Such a test would be of obvious value with blind subjects; moreover, it might complement the visual tests.

Method

Two 12-inch double-faced Vinylite records have been produced and are available to other investigators. The test records were played to individual patients in a room with only the examiner and the subject present. Instructions given the subject followed this general pattern:

> You are now going to hear a series of recorded sounds. Some of them will be familiar, some strange. Sounds have associations; they bring memories, stories, thoughts, and feelings to mind. This is true of music, cries, words, and so forth. You are going to hear 21 sounds, each lasting 20 to 50 seconds. [Most patients were not told the time in seconds but this is being done uniformly now.] Sounds have different meanings for different people. We are interested only in the meaning these sounds have for you. Tell me what your first thought is—what comes to your mind—what do the sounds tell you? If you wish, make up a story. Are you ready? Now, what does the sound mean to you?

All tests were administered by one of the authors (M. H.). The sounds were given one after another, usually without comment. If the tester felt that the subject was reluctant to reveal further thoughts, a carefully phrased query was made, usually by restating the last phrase the subject had used but in an inquiring tone, such as "Sounds like the train was going to leave?" As more experience was gained, the positions of examiner and subject were considered of some importance. The individual sat in a chair facing the loudspeaker and the tester sat behind the subject, out of sight, and took down the verbatim responses. Thus the effect of the appearance and attitude of the examiner was kept to a minimum.

The sounds employed are as follows:

1. Train (leaving station, crowd, depot noises).
2. Seagull, seawash background, conversation between two men:
    A. Is that enough . . . ?
    B. Yah.
    A. How much *is* that?
    B. (pause) Well, just about right . . . isn't it?
    A. I think that's just about right.
    B. Put that away now.
    A. Just about right.
    B. Would that cover it better?
    A. Put it away now; are you set?
    B. O.K., let's go.
3. Child cry and seagulls, soft background.
4. Clock striking and crowd noises: "Ya, ya."

---

[1] Read at the 108th Annual Meeting of the American Psychiatric Association, Atlantic City, May 12-16, 1952.

[2] From the Department of Neuropsychiatry, Stanford University Medical School. Supported by a grant from the Palo Alto Medical Research Foundation, Palo Alto, California.

5. Slow drip of water.
6. Organ and poem (male voice, clearly audible only on italicized words):

   To my years hath destiny denied
   The *glory of youth.* Ah,
   How art thou *fled* from view,
   Darling companion of my tender age,
   My hope, so *sorely* mourned!
   Is *this the fate* that time all men brings?
   Poor wretch, the *truth once plain*
   Thou wert *struck down with ease,* thy lifted hand
   Showing me cold death and a naked tomb
   Afar off grimly stand.
7. Cry for help and background noises.
8. Laughter and man and woman conversing:

   A. Where did that damn thing come from?
   B. What thing?
   A. Mother said you can't always get what you want in this world, sister.
   B. What did you expect?
   A. Where is Dad? Where did Dad go?
9. Music and sobs of woman (15 sec.) and man (15 sec.).
10. Man and woman arguing (man speaks in anger):

    Man: Now what did you do all that for today!
    (Mumbling answer).
    Huh? (Mumbling answer).
    You know what I'm talking about! (Mumbling answer).
    There's no excuse for it. (Mumbling answer).
    You know damn well what I mean! (Mumbling answer).
    You certainly did, and I'm ashamed. (Mumbling answer).
    Woman: Well, I wouldn't talk if I were you!
11. Music and water sounds, ending with rush of water (splash like fish in water or man drowning).
12. Father scolding child and child crying: Why! Why! Why!
13. Footsteps, faster, doors open, close, open, rush of footsteps.
14. Baby crying and woman singing softly.
15. Cry of woman and man.
16. Sucking sounds (human).
17. Groaning, rubbing, kissing (human).
18. Breaking wood, ripping cloth.
19. Drill.
20. Humming and whistling (man: Shut up!) more humming and whistling.
21. Train bells, (switch engine, loud bang, voices).

One hundred forty-six protocols form the basis of this study. The common factor in all but the college group is an enduring handicap. In the course of our study only three patients became sufficiently upset to necessitate discontinuing the test. Each of these was in a group other than those reported here; they were, respectively, a rheumatoid arthritic, and a depressed boy and woman. The constant recall of the crying, which seemed to be all they heard, was upsetting.

The samples are as follows:

Tuberculosis patients, who were tested at the San Mateo County Sanatorium, San Mateo, California, with cooperation of Dr. James Greenwell, Medical Superintendent. There were 46 patients, 25 males and 21 females, with an average age of 38.7 years and a range of 18-74 years. All of the patients were on strict bed rest or very limited exercise and had been in the hospital from one month to three years.

College students from Stanford University. There were 26 individuals, 15 males and 11 females, with an average age of 28.1 years and a range of 22-40 years.

Schizophrenic patients were tested at Agnews State Hospital, Agnew, California, with cooperation of Dr. Walter Rapaport, Medical Superintendent. They were all seriously ill committed patients. There were 23 paranoid patients, 8 catatonic, 7 hebephrenic, and 5 mixed type, a total of 43 subjects. Although the groups here were analyzed separately, for practical purposes and statistical reasons they will be considered as one group.

Blind children were tested at the California School for the Blind, Berkeley, with the cooperation of Dr. B. Lowenfeld. There were 21 children, 11 males and 10 females, with an average age of 14.5 years and a range of 10 to 16 years. A third of the children were totally blind or could only perceive light, while the others had vision of from 1/200 to 20/200. We have no other children to report at this time so the factors of blindness and childhood must be borne in mind when evaluating their responses in comparison with the other groups.

Amputees from the Korean War were tested at Letterman General Hospital, San Francisco, with cooperation of Col. James O. Gillespie, chief of professional services. Most of the patients were on crutches or in wheelchairs. There were only 10 patients in this group, so percentages will not be used. They had been hospitalized 3 to 12 months. Amputated limbs were as follows: one leg, 6; one foot, 1; both legs, 2; both arms, 1. For administrative purposes veterans were required to sign a statement stating in part: "I further agree to absolve the Government of the United States from any ill effects which may result in connection with said tests." This statement, implying alleged dangers, must have played a role in their attitudes and responses.

Table 1
Recognition of Selected Sound Components

Sound No.	T.B. $N=46$	College $N=26$	Schizophrenic $N=43$				Blind $N=21$
			Number				
	%	%	Paranoid	Other	Total	%	%
1. Train							
a. Crowd	50	77	8	8	16	37	15
b. Human	33	96	0	2	2	5	38
c. Street	50	69	8	2	10	23	48
2. Seawash, conversation							
a. Words repeated	17	27	9	13	22	51	48
b. Recognized background	15	54	0	0	0	0	81
c. Crime, police, etc.	26	55	4	4	8	19	90
3. Child cry, seagulls							
a. Jungle animals	13	54	5	4	9	21	67
b. Cry mentioned	37	54	14	7	21	49	57
c. More than one baby	48	23	5	3	8	19	52
4. Big Ben, crowd							
a. Bells (unspecified)	63	4	0	2	2	5	29
b. Church bells	46	12	2	9	11	26	24
c. Sadness, excitement, danger, etc.	52	31	6	4	10	23	28
5. Water drip							
a. Miscellaneous fluid	22	100	7	3	10	23	95
b. Running water	35	4	7	4	11	26	48
c. Poured into something	37	65	8	4	12	28	81
d. Poured out of something	15	15	8	2	10	23	0
6. Organ and poem							
a. Words recalled	4	38	1	0	1	2	5
b. Sadness—pleasure	15	80	1	10	11	27	57
c. Music	68	54	17	13	30	71	67
7. Cry for help, crowd							
a. "Help" repeated	57	35	14	10	24	57	62
b. Violence	41	62	7	4	11	26	95
c. Remote, unreal	57	35	3	1	4	9	24
8. Laughter and conversation							
a. Crowd mentioned	59	27	4	5	9	21	76
b. Laughter mentioned	52	50	10	6	16	38	71
c. Words repeated	20	8	6	4	10	24	24
9. Music, sobbing							
a. Crying mentioned	59	85	15	16	31	74	86
b. Woman crying	27	65	18	14	32	76	43
c. Unreal, remote	61	69	6	5	11	26	62
10. Man and woman arguing							
a. Man mentioned	95	92	18	11	29	69	86
b. Man at fault	27	12	0	0	0	0	14
c. Woman mentioned	89	85	13	9	22	52	90
d. Woman at fault	11	50	4	1	5	12	62
11. Music and water							
a. Water mentioned	42	81	14	10	24	60	100
b. Concert	63	27	8	10	18	45	14
c. Something in water	54	65	11	4	15	37	72
12. Father scolding son							
a. Provocation indicated	98	85	0	2	2	5	86
b. Reprimand (not spank)	70	23	11	8	19	49	43
c. Crying	20	30	6	4	10	26	57
d. Child's feelings	77	69	12	5	17	44	15

Table 1. (Continued)

Sound No.	T.B. $N=46$ %	College $N=26$ %	Schizophrenic $N=43$ Paranoid Number	Other Number	Total Number	%	Blind $N=21$ %
13. Footsteps and door							
a. Story indicated	67	42	12	3	15	38	25
b. Difficulty walking	29	8	1	0	1	3	15
c. Leaving	14	31	1	1	2	5	45
d. Fear, anxiety	24	50	5	1	6	15	55
14. Baby and woman singing							
a. Crying	52	54	11	13	24	62	55
b. "Mother"	10	85	11	13	24	62	60
c. Child's difficulty mentioned	69	73	1	7	8	20	50
d. Rocking, singing	67	35	12	14	26	67	60
15. Cry of man and woman							
a. Death, tragedy	44	32	5	1	6	15	20
b. Man and woman mentioned	54	40	9	2	11	28	46
c. Woman alone	28	8	7	12	19	49	7
d. Laughter (not on recording)	9	0	0	2	2	5	20
16. Sucking							
a. Total oral sounds	14	40	9	4	13	34	38
b. Person mentioned	33	56	7	1	8	22	28
c. Mechanical sounds	43	24	6	4	10	26	19
d. Sucking recognized	12	8	1	2	3	8	14
17. Groaning, rubbing, kissing							
a. Sleep described	60	16	5	6	11	27	11
b. Sleep mentioned	86	37	10	12	22	55	42
c. Dreaming	43	8	1	1	2	5	37
18. Breaking wood, tearing cloth							
a. Tearing	30	71	10	8	18	45	83
b. Wood mentioned	17	37	3	2	5	12	17
c. Grinding, drilling, machinery	32	0	0	3	3	7	0
19. Drill							
a. Saw, sawmill	59	61	4	3	7	20	13
b. Big log sawed	17	4	0	0	0	0	0
c. Power described	35	9	0	1	1	3	7
20. Humming, whistling, "shut up"							
a. Humming (woman)	71	48	12	8	20	53	63
b. Whistling (man)	64	48	12	11	23	61	58
c. "Shut up"	50	39	8	6	14	37	63
d. Man justified	19	26	3	0	3	8	42
21. Switch engine, loud "pop," crowd							
a. Train	60	17	9	9	18	49	33
b. Bell (train)	16	0	0	3	3	8	11
c. Gun or explosion	27	20	5	2	7	19	23
d. Streetcar	14	0	3	0	3	8	6

All patients were male, with an average age of 23.2 years and a range of 18-36 years. The small number of subjects, the single sex, and the signed statement set this group apart.

The responses of individuals to the sounds were analyzed according to their recognition of certain selected sound segments of each of the 21 sounds. For example, in the first sound, although there are a number of components, the three with the greatest response or divergence of response were selected for comparison and analysis in this study. The great difficulty in evaluating subjectively complex responses to subjectively complex auditory stimuli will be obvious. So as to avoid opinions and experience interpretations the sounds were fragmented into their component parts and the answers tabulated in relation to the frequency of similar responses to sound components.

The percentage of each of the groups recognizing or acknowledging in any way the component of each sound is given (Table 1). The number of amputees was too small to include.

The test required an average of about 35 minutes to administer, with a range of 25 to 85 minutes. The longest average time was 38.8 minutes in the case of the hebephrenic and catatonic schizophrenics, and shortest in the amputee group, 28 minutes. Eighty-six per cent of the subjects completed the entire test and 97% completed at least half. Only 7% rejected part of the test for personal reasons, viz., "I'm not nuts," etc., and 7% were unable to finish for reasons beyond their control, such as mealtime in the hospital, etc. Six patients refused the test completely before it was begun, three of these being amputees who had to sign a release.

Sample Responses

A 26-year-old paranoid schizophrenic patient responded to Sound 7 (cry for help) in the following manner:

> That would appear as if a person were calling for help, could be any state, either mental or physical...the person is calling from...I felt precisely that way a number of times, and I assure you it's not a very pleasant feeling. . .it's a feeling of extreme confusion. *Q*: Extreme confusion? *A*: Yes, or wanting something that was not available. There are certain personalities that affect you...that manner...having someone repeat the facts over and over...gave me that feeling of yelling for help. *Q*: Of yelling for help? *A*: Yes, that's in a remote sense. I want to be away from that person... I don't want to be with him. (25") In fact, I would say that sound would represent a mental picture of how I felt so many times. (15") I have an older brother ...and specially my older brother has always... well, not always, has at times (8") made more of an issue out of the things I said and developed them into problems. Whether he realized it or not, it bothered me...the fact that I had a paralytic stroke, having lived with him...after I left the University of California I went to live with him. I went to work in an oil well...I said I wanted a car...and he took it over from there (15") that we painted it and put a new top...and he took the car...it was a brotherly attitude on his part, but I don't believe he has much insight on my personal taste.

His responses in general revealed considerable fantasy, were frequently disjointed. He felt a strong identification in the cry for help; it connotes confusion and is, he believes, a reasonably good picture of his own mental state, while the crowd is being warned away from that person. There is no mention of help coming or of concern of the people for the plight of the individual. Although there are many words on the two records, the only one the patient recognized was the word "help."

In the protocol of an eleven-year-old blind boy his responses to the cry for help is as follows:

> Well, I have two things for that. Could be a boxer slowly getting knocked out . . . it could be in the old days, getting a beating, or being hanged. *Q*: Or being hanged? *A*: The way he stopped at the end, it sounded like he was being hanged... and a courtroom. A boxer would be saying "No, no." I guess that's it.

The patient does not mention the word "help," considers the man's plight could be one of two things, both violent physical punishment, and that it could have been in remote, ancient times. He seems to feel that the sudden end is the consequence of being choked, and imagines that the word "no" might be said by a boxer being beaten.

This protocol was one of rich fantasy, multiple associations employing "this or that" or "first this, then that." He often approached the sound as if he were attempting to classify it, then changed his mind in the course of his response. There was an unusual amount of identification of physical setting and time. His concern was largely with unseen dangers or with danger that would not directly affect him. Violence and danger were further indicated in the protocol by smuggling, robbing, shooting, hanging or being knocked down by an approaching dog.

He mentioned the words in five sounds. When he indicated food in Sound 10, it was being burnt (identifiable by nonvisual sense) and in Sound 13, the footsteps sequence, a burglar was robbing the house because everyone was asleep (eyes closed) and the burglar beat him over the head, knocking him out (so that his eyes would not open). The man who was sick was sleeping and "could be curtains."

These samples will serve to illustrate types of responses. Personal life associations and memories in responses were unusual and occurred most frequently in the college students and the schizophrenic patients, followed by the tubercular, blind, and amputees, in that order. There was an average of five personal associations in the college group, most frequently in Sound 1 (35%), and most frequently in the schizophrenics (17%) in Sound 9, music plus sobbing sequence. The paranoid schizophrenics and the other forms of schizophrenics have the same percentage of personal associations.

Results

A study of the responses of subject groups can be summarized as follows:

Tuberculosis patients. Heard appeal most often, cries least often. Angry crowd, death and tragedy (mood) but not violence (action) were inferred most often. They heard laughter most often, even when not in actual recording. They were keenest to walking difficulties but least keen to perceive oral sounds. Though they described hungry babies crying, like all other groups, the children were never fed, and only once "going to be fed soon."

College students. It can be surmised that they had emphatic affective responses, with strong feelings for right and wrong, with considerable anxiety. They had a high response to crying and a low one to laughter. Their responses indicating violence and sadistic behavior were equalled by those of the blind children. One does not know the incidence of neurosis among this sample, but this could be a significant factor. They heard the doors open and close most often; like all groups, if the doors were heard only to open or only to close, the males predominated over the females about 2:1. If doors both opened and closed, males predominated about 4:3, which is not considered significant because of the slight male predominance of the total sample. In the tuberculosis group, for example, it was only males who heard the door open, while in the college group the two female subjects indicating the door opening only mentioned that it opened two or three times.

Schizophrenic patients. The paranoids paid little attention to words; violence, death, tragedy, and laughter were least frequently heard, but crying commonly. Their responses were definitive, not often guarded, were rich but with little movement, oral sounds or needs or gratification responses.

Blind children. These found the elements of violence, death, and struggle more frequently than any other group. They detected background noises keenly but tended to reject other components of sounds when they fixed their verbalization on the danger elements. They were inclined to refer to the woman and her efforts most often. Their own handicap is reflected by comments about walking, unseen danger, and warning signals. They referred to sucking most often but not to sleep unless there was danger in the sounds.

Considering the small sample of amputees the most striking feature was the poverty of responses and the subjects' expressed difficulties in understanding the sounds. They practically never referred to words but frequently referred to the affective sounds of laughing and crying. When music was associated with crying, music was rarely heard; on the other hand, when music was heard with water, music was frequently mentioned. It would seem the impact of crying tended to overshadow music. They referred to violence only once. It is interesting that in the final sound, where a gunshot or explosion is often inferred, this was only once acknowledged, then as "pop." These veterans, having returned from the Korean War within recent months with loss of limbs, excluded the verbalization (and recognition?) of the significant meaning of violence. They were most productive of feeling responses in the sound of the father scolding the child.

Discussion

There are certain limitations to the present study. Although we analyzed 146 protocols, we do not yet know the value of this test as a projective technique, nor its predictive or interpretive importance. We do not have a group of nonblind children, and, in essence, our study of blind children permits only comparison of them to sighted adults. This is significant, as it is the setting in which all psychotherapy with the blind takes place. Subsequent study of the nonblind will be reported. Our so-called normal group (college students) probably constitutes a segment of presumably non-ill young adults. It is considered as a group for comparison rather than controls, which must include subjects of all ages and in all walks of life. This distinction between normal controls and sample for comparison is important.

Despite the fact that we do not yet know the value of this test as a projective technique, it meets the criteria for such a test as set forth by Rapaport [3]: (*a*) economical, i.e., not time-consuming; (*b*) simple, i.e., feasible for office examination and not requiring complicated or bulky apparatus (but does require a record player of good tone quality); (*c*) impersonal, i.e., does not involve the need to build up time-consuming transference relations or complexes of intense interpersonal relationships, such as are present in psychodrama; (*d*) limited to one segment of "behavior"; (*e*) standard in the sense that all subjects face the same material and the same degree of cooperation from the examiner.

Summary

A sound projection test employing 21 sounds on two 12-inch double-faced Vinylite records is described. They include simple and complex mixtures of verbal, nonverbal, mechanical, and natural sounds. Copies of these records are available for research purposes.

Protocols of 146 subjects are analyzed. These include tuberculosis patients, schizophrenics, college students, blind children, and Korean War amputees. No norms can be given. The value of the test in interpretation of personality structure remains to be understood.

Tuberculosis patients heard words of appeal and crying and inferred violent action more than other groups. They described the mood of sadness and tragedy as well as laughter more than the others.

College students gave a high response to crying but low to laughter. They indicated a high degree of violence and sadistic behavior responses. They gave the most frequent human and mechanical movement responses.

Schizophrenic patients ignored the words and referred least often to violence, death, and tragedy as well as to laughter. They acknowledged the crying in large proportion.

Korean War amputees rejected the explosive and gun responses and the violent sounds, but gave rich associations with the father-scolding-son sound.

The blind children gave the most frequent recognition of violence and sadistic activity, heard the words of appeal and anger most often, and reflected their own handicap in their responses of unseen danger.

Received December 22, 1952.

References

1. Skinner, B. F. The verbal summator and the method for the study of latent speech. *J. Psychol.,* 1936, **2**, 71-107.
2. Stone, D. R. A recorded auditory apperception test as a new projective technique. *J. Psychol.,* 1950, **29**, 349-353.
3. Rapaport, D., Gill, M., & Schafer, R. *Diagnostic psychological testing.* Chicago: Yearbook Publishers, 1945. 2 vols.
4. Wilmer, H. A. An auditory sound association technique. *Science,* 1951, **114**, 621-622.
5. Wilmer, H. A., & Husni, May. An auditory projection technique for psychiatric research. *J. Lancet,* 1952, **72**, 17-24.

Journal of Consulting Psychology
Vol. 17, No. 5, 1953

Overinclusive Thinking in a Schizophrenic and a Control Group[1]

Seymour Epstein
University of Massachusetts

For any task to be properly performed, it is necessary to include behavior appropriately related to the task and exclude behavior inappropriately related to it. A failure adequately to include or exclude is held by Cameron [1, 3, 4, 6, 7, 8] to be characteristic of certain types of maladjustment. According to Cameron overinclusion favors disorganization whereas underinclusion[2] is resorted to in an attempt to preserve or achieve stability. In terms of diagnostic categories, the conversion-hysteric is viewed as a pathological underincluder, the schizophrenic as a pathological overincluder. A review of the literature reveals several studies which have reported overinclusive thinking among schizophrenics [1, 2, 3, 4, 14, 15]. Some have noted underinclusive thinking as well [14, 15]. The purpose of the present investigation is to investigate more thoroughly schizophrenic overinclusion by attempting to answer the following questions:

1. How prevalent is overinclusive thinking among schizophrenics? Do they overinclude more than matched controls? Is overinclusion more characteristic of some schizophrenic subtypes than others?

2. Is overinclusive thinking more characteristic of schizophrenics than underinclusive thinking?

3. What variables, such as sex, vocabulary level, abstract-thinking ability, and intellectual impairment, are related to verbal overinclusion?

4. What kinds of words are schizophrenics most apt to overinclude? Is overinclusion a general characteristic of schizophrenic thinking or is it restricted to certain kinds of stimuli having particular significance in terms of individual needs and expectancies?

Procedure

Description of the Inclusion Test. For the purpose of this investigation an objective paper-and-pencil test, the Inclusion Test, was constructed.[3]

The test is composed of 50 items, each consisting of a key word followed by five response words and the word "none." The testee is instructed to underline all response words which designate things or concepts required for the complete thing described by the key word. An example of an item is as follows:

MAN arms shoes hat toes head none

A correct performance requires underlining the response words *arms, toes,* and *head,* as every complete man must have, among other things, arms, toes, and a head. For each correct choice omitted, a score of one underinclusion is received; for each incorrect choice selected, a score of one overinclusion is obtained. Many of the items contain words only superficially related to the key word, such as klang associations. Some items contain words meant to be emotionally charged. Several key and response words are neologisms, and others are so ambiguous that they cannot be meaningfully scored. This latter type is intended to induce confusion by forcing impossible discriminations.

---

[1] Based upon a thesis submitted in 1950 in partial fulfillment of the degree of Master of Arts at the University of Wisconsin. The writer wishes to express his gratitude to Dr. Ann Magaret for her advice and guidance.

An abstract of this paper was presented at the 1951 meeting of the Midwest. Psychol. Assn.

At the time this study was carried out the author was a trainee in the Veterans Administration program. He wishes to express his appreciation to this organization for its support of the study.

[2] Underinclusion, as used here, has the same meaning as Cameron's term, overexclusion, and has been substituted merely for reasons of readability.

[3] The scoring system for the Inclusion Test was determined by an item analysis of data obtained from a preliminary administration to 61 college students. In its final form the college group made an equal number of errors of overinclusion and underinclusion.

Subjects. The patient sample consisted of 38 schizophrenics,[4] 19 of whom were males. Only those schizophrenics were selected who had not experienced a shock treatment for at least two weeks, who were within the age range 20-42, in whose case record no question regarding correctness of diagnosis had ever been raised, and who were assigned to the less disturbed wards.

The control group consisted of 20 men and 25 women. They were selected from a larger group of 97 tested, on the basis of providing a close match with the schizophrenics on age and vocabulary level. The controls represented a wide range of occupational backgrounds, having been solicited from among firemen, recreation groups, housewives, hospital attendants, secretaries, and professional people.

Testing procedure. The Shipley–Institute of Living Scale [13] was administered together with the Inclusion Test in order to allow for an evaluation of the relationship between its scores with overinclusion and to provide a means for matching schizophrenics and controls on vocabulary level. Subjects were tested individually or in small groups. Every effort was made to put the subject at ease and to insure his comprehension of the task before he was allowed to begin. Total testing time rarely exceeded three-quarters of an hour.

Table 1

Analysis of Variance for Overinclusion Scores of 38 Schizophrenics and 45 Controls

Group	Mean	Sum of squares	*df*	Mean square	
Normal	12.49	1464.87	1	1464.87	12.03***
vs.					
Schizophrenic	20.92				
Male	15.44	61.39	1	61.39	1/1.98
vs.					
Female	17.16				
†Vocab. 1	19.62	328.65	2	164.32	1.35
vs.					
Vocab. 2	14.50				
vs.					
Vocab. 3	15.85				
(Error)		9493.96	78	121.72	

*** Significant beyond the .001 level of confidence.

† Vocab. 1 consists of scores of 21-25 vocabulary items correct; Vocab. 2 of 26-29 correct; Vocab. 3 of 30-40 correct.

Results

Results are presented in the order of the questions raised at the beginning of the paper.

Prevalence of overinclusion in schizophrenia. The main object of this study was to determine whether overinclusion in a schizophrenic group, as a whole, occurs more frequently than in a control group. The results indicated in Table 1 suggests that it does. The mean overinclusion score for the schizophrenics is nearly double that for the controls. The difference between group means is significant beyond the .001 level of confidence.

The next question to be considered is what percentage of the schizophrenics can be classified as overincluders. This is a difficult matter to decide upon because of a lack of normative data defining overinclusion. However, inspection of the distributions suggests that a score of 18 or more overinclusions might be used to determine pathological overinclusion, as such a cutting point would separate the mode of the schizophrenic distribution from the mode of the controls. Sixty-three per cent of the schizophrenics and 22 per cent of the controls fall beyond this point. Thus, many schizophrenics (37%) do not overinclude more than most controls, and overlap between the groups rather than sharp separation is the case.

An analysis of variance did not indicate differences in overinclusion among the subtypes of schizophrenia. The sample included 7 "mixed" or "undetermined," 12 "hebephrenic," 14 "paranoid," 4 "catatonic," and 1 "simple." Despite the lack of significant differences, the order of the means of the subtypes did suggest that the greater the personality disorganization, the greater the overinclusion.

Overinclusion in relation to underinclusion. An analysis of variance for the underinclusion score, analogous to he one for overinclusion in Table 1, reveals no difference betwen the schizophrenic and control group in mean underinclusion. Of interest is the observation that the controls make about an equal number of erros of overinclusion and underinclusion, with means respectively of 12.49 and 12.10, while the schizophrenics make almost twice as many overinclusions as underinclusions, with means respectively of 20.92 and 12.20. It thus appears that the difficulty schizophrenics have in verbally establishing boundaries in the Inclusion Test is predominantly a result of a tendency to overinclude.

---

[4] The writer wishes to express his appreciation to the staff of Mendota State Hospital, particularly to psychologist Phyllis Earnest, for cooperation in securing the patient sample.

Overinclusion in relation to other variables. Inspection of Table 1 reveals that, within the range of the sample tested, neither men nor women have an option on overinclusion, nor is it influenced by vocabulary level. Abstract-thinking ability, as measured by the Shipley scale, is also not found to be significantly correlated with overinclusion.[5] On the other hand, the product-moment correlation coefficient of overinclusion and the Shipley measure of "intellectual impairment" is .37 for the schizophrenic group ($N=38$) which, when evaluated by Fisher's t, is significantly greater than zero at beyond the .05 level of confidence. This result is in line with the interpretation of the order of the means of the subtypes of schizophrenia, i.e., the greater the disorganization, the greater the overinclusion.

Kinds of materials schizophrenics overinclude. A chi-square analysis reveals that 10 of the 50 items of the Inclusion Test are handled significantly differently by schizophrenics and controls at beyond the .05 level of confidence. This is considerably beyond the chance expectancy of 2.5 items. All five items in the test that have neologistic key words are represented in the significant 10. In general, controls make no choice or one choice following a neologism, while schizophrenics not infrequently make two. As controls and schizophrenics were matched on vocabulary, their differential handling of neologisms cannot be explained in terms of familiarity with words. A possible explanation was suggested by spontaneous comments made by some of the patients, indicating an uncertainty about what they did and did not know. For example, one patient, upon coming to the neologism "to-pitch," remarked, "Isn't it a shame the way words you've used all your life suddenly escape you?"

Inspection further indicated that the schizophrenics inappropriately chose words related to material things and sense impressions. Thus, more schizophrenics than controls incorrectly selected *wood* as necessary to a *Box,* whereas more controls than schizophrenics incorrectly selected *corners.* Also, schizophrenics, more so than controls, tended to select specific examples of a category as necessary for that category (i.e., a *lion* is required in order to have an *Animal*). Such behavior is in line with Goldstein's [11] description of schizophrenic thinking as concrete and indicative of a "loss in the abstract attitude."

Qualitative analysis further suggested that the schizophrenics, more than the controls, responded with "loose associations" rather than with meaning. By this we mean they chose words only incidentally related to the key word, such as is the case with klang associations. As the Inclusion Test contains a great number of associationistic choices, this type of response contributed heavily to the schizophrenic overinclusion score.

That schizophrenics respond to neologisms and choose response words which are concrete or only superficially and associationistically related to the key word suggests that overinclusion is a general characteristic of schizophrenic thinking and does not occur only in areas having particular emotional significance to the individual.

Discussion

The findings in this study support Cameron's view that schizophrenic thinking is overinclusive. Why no underinclusion was found raises a question, as both Shneidman [14] and Zaslow [15] reported underinclusive as well as overinclusive behavior in their groups. Whereas other explanations are possible, we suspect the discrepancy in findings is a result of differences in the tasks in two respects. For one, the Inclusion Test offers many associationistic choices, which the schizophrenic seems particularly apt to respond to. Thus, overinclusion is favored by the nature of many of the individual items. For another, the test as a whole is not rigidly structured in that it offers many bizarre choices for which there are no correct responses. It may be that underinclusion, when it occurs among schizophrenics, is a reaction formation to a more basic overinclusive tendency in an attempt to prevent or ward off disorganization, and the Inclusion Test, because of its confusing nature, disrupts the development of a rigid defense system. Of interest, in this connection, was the reaction of one patient who underlined only the word *none,* having explained that no response, including the ones indicated in the illustrative examples at the beginning of the test, met her criteria of perfection. She was negativistic throughout the session, becoming increasingly irked with the test, and finally refusing to complete it. Had she proceeded in her test behavior to the end, she would have received a score of extreme underinclusion. One is tempted to conjecture here that she was able to prevent the collapse of her underinclusive defense system only by withdrawing from the task. That underinclusion may be a reaction formation to overinclusion is also in line with reports on the use of rigidity as a defense mechanism [9, 10, 11].

---

[5] The absence of a significant correlation cannot readily be attributed to unreliability of the Inclusion Test, as the corrected split-half reliability coefficient for the overinclusion score was found to be .90 ($N=83$).

Why the schizophrenic overincludes, associationistically or otherwise, is not immediately apparent. Cameron, who originally used the concept of overinclusion descriptively [4], has more recently attempted to explain it in the following manner: ". . . a strong hypothesis, arising in response to anxiety-provoking uncertainty, begins by broadening the generalization gradient and developing overinclusion. . ." [8, p. 306]. Whereas this explanation may account for certain kinds of overinclusion, it does not account for the results observed in this investigation, for on the basis of a "strong hypothesis" theory one would predict underinclusion as well as overinclusion in the test used here. It might simplify matters if two types of overinclusion were differentiated. One, the kind resulting from strong needs or expectancies, could best be included under the concept of selective perception, and would involve overresponding to material related to subjective hypotheses and simultaneously underresponding to other material. This investigation has revealed a different type of overinclusion, which we suspect is a result of a failure in discrimination, arising out of a defect in attending. That schizophrenics suffer from an attention defect has been shown in several studies [5, 12]. Whether such a defect impairs discrimination and favors diffuse generalization of the type demonstrated by our schizophrenics remains for future research to determine. Operationally, this hypothesis can be tested by determining whether overinclusion increases as attention level is experimentally decreased, and whether it does so more than underinclusion. Obviously, the prediction of increasing overinclusion with decreasing attention can be expected to hold only within limits, for with extreme inattention the individual would not be responding at all to the task at hand, and so, in a sense, would be underincluding. It is therefore hypothesized that the curve of overinclusion as a function of decreasing attention rises gradually at first, then falls sharply when almost all attention is withdrawn from the task at hand.

This study raises several questions. For one, it has been shown that on the basis of Cameron's concept of overinclusion the reaction of schizophrenics to a particular test situation could be predicted. But the schizophrenic represents only the overinclusive end of the continuum. Is Cameron's theory correct in viewing the hysteric as an underincluder, and would such behavior be opposite but analogous to schizophrenic overincluding?[6] Another question that arises is whether overinclusion is characteristic only of schizophrenic behavior, or whether it is equally prominent in other psychoses. We believe that any state of disorganization favors overinclusion, but this remains to be tested. A third matter that might profitably be investigated is what other correlates of overinclusion or underinclusion, besides "intellectual impairment," can be found to help relate overinclusion and underinclusion to clinical behavior. For example, how does the non-overincluding schizophrenic differ from the overincluding schizophrenic? Does he have a more favorable prognosis? Do rigidity and defensiveness favor underinclusion, as was hypothesized? Can certain kinds of hallucinations be explained in terms of the overinclusion of normally unheeded perceptions? A positive relationship between some kinds of hallucinatory experience and score on the Inclusion Test would support such a view.

---

[6] In a preliminary study the Inclusion Test was administered to six patients demonstrating uncomplicated conversion-hysteric symptoms. Their mean scores and the control group's for both overinclusion and underinclusion were almost identical. If this result is confirmed by larger studies, it would suggest that the hysteric does not underinclude extensively, but perhaps only in regard to material dynamically related to his particular symptom. Such behavior would not be opposite and analogous to the overinclusion of the schizophrenic, which is diffuse.

Summary

In order to investigate overinclusive thinking among schizophrenics an objective paper-and-pencil test, the Inclusion Test, was constructed. The Inclusion Test yields a score of overinclusion and a score of underinclusion. A comparison on this test of matched groups of 38 schizophrenics and 45 controls resulted in the following findings:

1. Schizophrenics, as a group, overinclude significantly more than controls. Whereas controls, on an average, make about an equal number of errors of overinclusion and underinclusion, schizophrenics average about twice as many errors of overinclusion as underinclusion. However, overinclusion is not characteristic of every schizophrenic.

2. Schizophrenics, as a group, do not behave differently from controls in regard to underinclusion.

3. Sex, subtype of schizophrenia, vocabulary level, and abstract reasoning are not related to degree of overinclusion. On the other hand, "intellectual impairment" is positively and significantly correlated with overinclusion.

4. Overinclusion does not appear to be confined to material with apparent strong personal significance for the schizophrenic. Rather, the schizophrenic tends to overinclude stimuli which are but associationistically or concretely related to the task at hand. He even overresponds to words the meaning of which he does not know.

In order to explain some of the above findings it was hypothesized that underinclusion, when it is found among schizophrenics, is a reaction to a more basic overinclusive tendency in an attempt to prevent disorganization. It was further suggested that the kind of overinclusion observed arises from a defect in focusing attention.

Received January 13, 1953.

References

1. Cameron, N. Reasoning, regression and communication in schizophrenics. *Psychol. Monogr.*, 1938, **50**, No. 1 (Whole No. 221).
2. Cameron, N. A study of thinking in senile deterioration and schizophrenic disorganization. *Amer. J. Psychol.*, 1938, **51**, 650–665.
3. Cameron, N. Deterioration and regression in schizophrenic thinking. *J. abnorm. soc. Psychol.*, 1939, **34**, 265–270.
4. Cameron, N. Schizophrenic thinking in a problem-solving situation. *J. ment. Sci.*, 1939, **85**, 1012–1035.
5. Cameron, N. The functional psychoses. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 861–921.
6. Cameron, N. Experimental analysis of schizophrenic thinking. In J. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley: Univer. of Calif. Press, 1946. Pp. 50–63.
7. Cameron, N. *The psychology of behavior disorders — a biosocial interpretation.* Boston: Houghton Mifflin, 1947.
8. Cameron, N. Perceptual organization and behavior pathology. In R. R. Blake, & G. V. Ramsey (Eds.), *Perception—an approach to personality*. New York: Ronald, 1951. Pp. 283–306.
9. Fisher, S. Patterns of personality rigidity and some of their determinants. *Psychol. Monogr.*, 1950, **64**, No. 1 (Whole No. 307).
10. Frenkel-Brunswik, Else. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.*, 1949, **18**, 108-143.
11. Goldstein, K. Methodological approach to the study of schizophrenic thought disorder. In J. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley: Univer. of Calif. Press, 1946. Pp. 17-39.
12. Hunt, J. McV., & Cofer, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 971–1032.
13. Shipley, W. S. *Shipley–Institute of Living Scale for Measuring Intellectual Impairment.* Hartford, Conn.: Institute of Living, 1940.
14. Shneidman, E. S. Schizophrenia and the MAPS test: a study of certain formal psychosocial aspects of fantasy production in schizophrenia as revealed by performance on the Make a Picture Story (MAPS) test. *Genet. Psychol. Monogr.*, 1948, **38**, 145–223.
15. Zaslow, R. W. A new approach to the problem of conceptual thinking in schizophrenia. *J. consult. Psychol.*, 1950, **14**, 335–339.

Journal of Consulting Psychology
Vol. 17, No. 5, 1953

Intellectual Deficit in Organics and Schizophrenics

Sheldon R. Rappaport

Department of Psychiatry, Albert Einstein Medical Center, Southern Division, Philadelphia, Pa.

This study, part of which was reported in a previous paper [20], is an investigation of the role of behavioral accessibility in intellectual deficit. This paper specifically deals with determining whether organic or schizophrenic psychotics show significantly greater intellectual deficit on specific tests and whether such deficit can be attributed to organic or functional factors, the latter rated by behavioral accessibility scores on the Elgin Test Reaction Scale [20].

The previously reported selection of the sample and conditions of the study are summarized below. The test batteries selected to be administered to these patients were comprised of test items believed to reflect most adequately intellectual areas in which psychotic patients may show deficit. Vocabulary was also measured because it is widely held as the best estimate of premorbid level of function. It is recognized that the designations for these areas are merely summative indices of more complex functions [17].

Because of the difficulty of maintaining the cooperation and attention of psychotics, it was decided that the selected tests should be arranged in two batteries, each requiring approximately the same time (usually around 40 minutes) for administration. To introduce each battery with a test which would facilitate rapport, to permit proper placement of the digit-span test, and to allow for delayed recalls, the following order of tests was devised:

Test I

1. *Information,* subtest of the Wechsler-Bellevue, I [26]; an index to availability in consciousness of a fund of general information obtained by cumulative interest in the environment.

2. *Recall I,* the news item from X, 3 of the Stanford-Binet, L [22]; an index to the accuracy of immediate recall of meaningful material.

3. (*a*) *Digits Forward* and (*b*) *Digits Reversed,* subtest of the Wechsler-Bellevue, I; an index to auditory attention span.

4. *Memory-for-Designs test,* by Graham and Kendall [7, 8]; an index to the accuracy of visual organization and reproduction of imagery.

5. *Recall II* (same as 2 above); an index to the accuracy of retention and delayed recall of meaningful material.

6. *Block Designs,* subtest of the Wechsler-Bellevue, I; an index to analytic-synthetic thinking.

7. *Vocabulary,* the Thorndike short-form of the Binet Form L Vocabulary [24]; the best single estimate of premorbid level of intellectual function.

Test II

1. *Comprehension,* subtest of the Wechsler-Bellevue, I; an index to understanding of social factors and requirements, as fundamental to social judgment.

2. *Word Pairs I,* the word pairs from the Hunt-Minnesota Test of Organic Brain Damage [10]; an index to formation of new, rote associations.

3. *Vocabulary,* subtest of the Wechsler-Bellevue, I; the best single estimate of premorbid level of intellectual function.

4. *Sorting test,* the Weigl-Goldstein-Scheerer Color-Form Sorting Test [6]; an index to concept formation.

5. *Word Pairs II* (same as 2 above); an index to retention and delayed recall of newly formed rote associations.

6. *Cube test,* the Goldstein-Scheerer Cube Test [6]; an indication of ability to profit from concrete aids in the performance of analytic-synthetic thinking.

Tests I and II were administered to a patient at the same hour on two successive days, according to the prescribed procedures for each of the tests used. At each of these two sessions the patient's behavioral accessibility was rated on the Elgin Test Reaction Scale (TRS). Those patients who failed any of the first four block designs in Test I were given the Cube test, designs I-VI, at the end of Test II.

Scoring Procedure

Wherever established scoring procedures were available, they were followed. Two tests, the Sorting test and the Cube test, had no such standardized, quantitative scoring methods.

To obtain a quantitative means of evaluating the Sorting test, the writer devised a system of weights for the various performances on this test. These are described in Table 1. This scale is based primarily upon the principles set forth by Goldstein and Scheerer [6].

Table 1

Scale of Weights for Performance on the Weigl-Goldstein-Scheerer Color-Form Sorting Test

	Spontaneous sorting and shifting, both concepts abstract
19	Both verbalized
18	One verbalized
17	Neither verbalized
	Spontaneous sorting and shifting, one concept abstract
16	Both verbalized
15	One verbalized
14	Neither verbalized
	Spontaneous sorting and shifting, both concepts concrete
13	Both verbalized
12	One verbalized
11	Neither verbalized
	Induced sorting and shifting, both concepts abstract
10	Both verbalized
9	One verbalized
8	Neither verbalized
	Induced sorting and shifting, one concept abstract
7	Both verbalized
6	One verbalized
5	Neither verbalized
	Induced sorting and shifting, both concepts concrete
4	Both verbalized
3	One verbalized
2	Neither verbalized
1	Failed

To quantify the six steps involved in the Cube test [6], Boyd's method was used [4]. Accordingly, making a correct pattern and giving a correct explanation of this on step 1 are each given a weight of 5 points, or a total weight of 10 points. Attaining a correct pattern and explanation on step 2 is given a total score of 5 points; on step 3, 4 points; on step 4, 3 points; on step 5, 2 points; on step 6, 1 point. Hence, because the first 6 designs were used in this study, a total score of 60 was possible on the Cube test.

Analysis of the Data

The schizophrenics and organics were combined and then were divided into thirds according to the magnitude of their TRS scores. The lowest third of the group had TRS scores below 67, the middle third had scores between 67 and 74, and the upper third had scores of 75 or more. These scores were used to differentiate both the schizophrenics and organics into thirds. Which third a patient was in depended on his TRS score for that day's testing. Hence, because of change in behavior, a patient might be in the upper third for Test I and in the lower third for Test II.

An analysis of variance was made for 14 of the 15 tests comprising Test I and Test II. For each test, estimates of the population variance were obtained from variance attributable to diagnosis, level of behavioral accessibility, the interaction of these two components, and from the within-group variance. The Cube test was not so treated because not all patients were to receive it. On this test, passing was set arbitrarily at achieving a score of 3 on each of the six designs, or a total score of 18. However, if a total score of 18 was achieved without making a score of 3 or more on each design, failure was recorded. Such inconsistency as achieving a score of 10 on two of the designs and then failing the others was not considered manifestation of having learned from the Cube test aids. The scores on the Cube test were treated by determining the critical ratio between the percentage of schizophrenics and organics who failed any of the first four designs of the Block Design test and passed the Cube test.

Results

An analysis of variance was made to determine if test scores would be differentiated by diagnosis, by levels of behavioral accessibility, or by the interaction of these. The relationship between mean scores of schizophrenics and organics at each level of accessibility is shown in Figures 1 and 2.[1] The mean scores and the

---

[1] It should be remembered that the score values hold different significances for the various subtests, with the exception of those reflecting the Wechsler-Bellevue weighted scores. Except in the case of the latter scores Figures 1 and 2 are meaningful only in comparison of the relative position of the scores of organics and schizophrenics within a subtest.

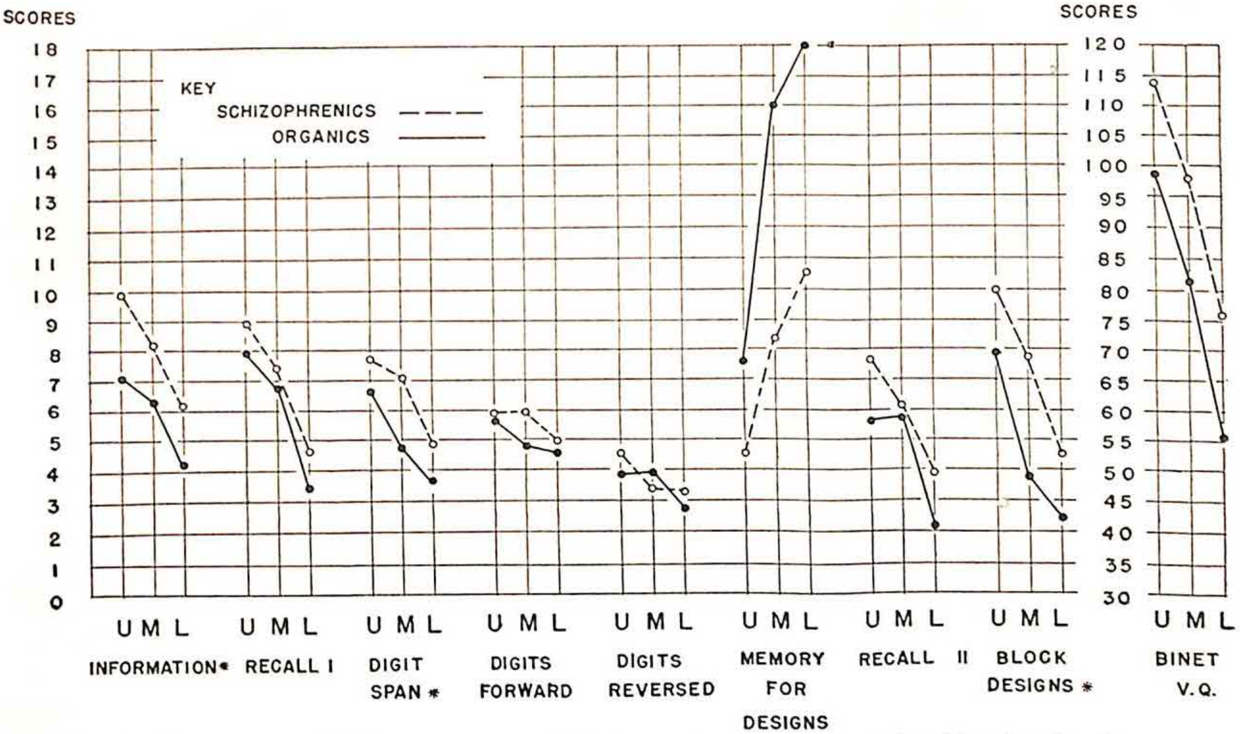


Fig. 1. Mean scores (*indicates mean weighted score) of organics and schizophrenics for upper, middle, and lowest thirds of behavorial accessibility on subtests of Test I.

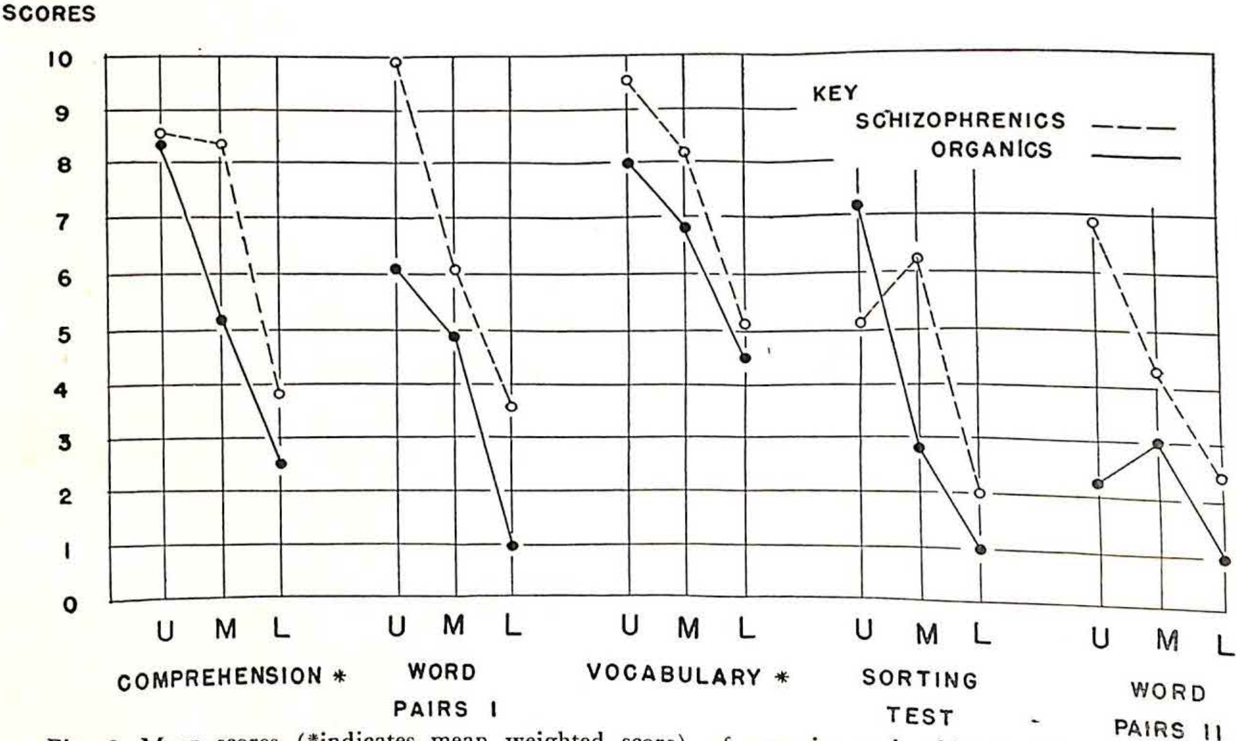


Fig. 2. Mean scores (*indicates mean weighted score) of organics and schizophrenics for upper, middle, and lowest thirds of behavorial accessibility on subtests of Test II.

Table 2

F Tests of Mean Scores on Tests I and II by Levels of Accessibility and by Diagnosis

	Levels of accessibility Means Up. 3rd	Mid. 3rd	Low. 3rd	% L of C for F	Diagnosis Means Schiz.	Organic	% L of C for F
Test I							
Information*	8.85	7.35	5.68	1	7.97	6.12	1
Recall I	8.39	7.08	4.34	1	6.79	6.39	<5
Digit Span*	7.29	6.03	4.58	1	6.44	5.17	1
Digits Forward	5.80	5.45	4.84	1	5.56	5.02	1
Digits Reversed	4.27	3.70	3.13	1	3.85	3.46	<5
Memory-for-Designs	5.68	11.60	12.29	1	7.88	13.39	1
Recall II	6.93	5.95	3.50	1	5.77	5.00	<1
Block Designs*	9.24	6.10	4.18	1	7.37	5.05	1
Binet VQ	108.76	91.63	72.05	1	95.62	83.02	1
Test II							
Comprehension*	8.56	7.00	3.59	1	7.12	5.63	1
Word Pairs I	8.21	5.59	3.07	1	6.56	4.13	1
Vocabulary*	9.15	7.61	5.03	1	7.73	6.83	5
Sorting Test	5.76	4.76	1.83	5	4.66	3.68	<5
Word Pairs II	5.53	3.80	2.17	<5	4.62	2.50	<5

* Means derived from Wechsler-Bellevue, I weighted scores.

Table 3

Significant t's for Scores on Tests I and II between Diagnoses and among Levels of Behavioral Accessibility

	Between schiz. & org. U.	M.	L.	Among schizophrenics U-M	M-L	U-L	Among organics U-M	M-L	U-L
Test I									
Information	3.40**	2.26*	1.97*	2.40**	2.80**	5.43**		1.99*	2.65**
Recall I					3.02**	3.56**		2.45**	3.19**
Digit span		2.45**			2.62**	3.59**			2.38**
Digits f'wd.		2.86**			2.82**	2.80**			
Digits rev.					1.97*	3.71**			2.24*
Mem.-for-designs		3.63**	2.93**			3.29**	3.57**		3.69**
Recall II					2.11*	3.79**		2.41*	2.29*
Block designs	2.40**	4.71**	2.10*	2.96**	4.16**	7.40**	4.36**		4.79**
Binet VQ	2.15*	2.20*	2.44**	2.60**	3.50**	6.38**	2.11*	2.93**	4.64**
Test II									
Comprehension		4.13**			5.95**	4.86**	3.01**	2.17*	4.11**
Word pairs I	2.31*			3.40**	2.19*	5.01**		2.00*	2.33*
Vocabulary				2.05*	4.34**	6.04**			2.53**
Sorting test		2.25*			2.61**				2.05*
Word pairs II	4.54**			2.96**	2.39**	5.29**			

* 1% level of confidence;
** 5 to 2% level of confidence.

levels of confidence of F are in Table 2. These F tests were derived from the ratio of the variance of behavioral accessibility to the within-class variance and from the ratio of the variance of diagnosis to the within-class variance. This procedure was followed because in each case the interaction variance was found not to be significantly larger than the within-class variance [14]. Exception to this was found in Delayed Word Pairs.[2]

---

[2] Here the interaction variance was larger than the within-class variance at the 5% level of confidence and thus was used as the error variance. Neither of the F's was then significant. However,

Because the F tests showed significant differences in scores among levels of behavioral accessibility and between diagnoses, t tests[3] were employed to ascertain where the differences lay. The magnitudes of the significant t's and their corresponding levels of confidence are shown in Table 3.

Of the organics who failed the Block Designs, 60.60% passed the Cube test, while this was true for only 53.40% of the schizophrenics. The critical ratio between these percentages was 7.69, highly significant.

Discussion of Results

Inspection of the preceding tables and graphs indicates that on all subtests the schizophrenics' scores vary directly with their level of accessibility. With the exception of two subtests, this is also true of the organics' scores. These data serve to substantiate the high degree of relationship already established between psychotics' test scores and their behavioral accessibility [20]. These results further suggest that regardless of the ultimate basis for the illness, psychotics show a disturbance of thought process that has at its cause a large behavioral or functional element. This agrees with the conclusions of Rapaport [19] and others in contraindicating the traditional psychiatric viewpoint of a dichotomy of mental disorders into those of neurological and those of psychological origin.

Indeed, scores on Recall I, Digits Reversed, Recall II, and Wechsler-Bellevue Vocabulary seem to reflect only behavioral accessibility and do not differentiate organics from schizophrenics at all. This would suggest that findings of vulnerability of, for example, digit span [1, 2, 15] in organics have been due to behavioral inaccessibility rather than specifically to brain damage. This further suggests that vocabulary cannot be regarded as resistant to encroachment by disease process and therefore an absolute index to premorbid level of function when vocabulary is so vulnerable to behavioral inaccessibility (viz. Fig. 1). Despite the intimation of Allen [1, 2] and others that behavior plays a role in intellectual impairment, there still seems to be too great a proneness to interpret "organic" or "deterioration" tests as evidence of immutable deficit in attention, retention, etc., due solely to brain damage.

On the other hand, the scores on Information, Block Designs, and Thorndike-Binet Vocabulary are significantly higher for the schizophrenics than for the organics. Although these results agree with previous studies [1, 2, 13] in that Block Designs shows greater deficit in organics than does Information or Vocabulary (Table 2), the present study indicates that when compared with schizophrenics, the organics show deficit on all three of these tests. This cannot be accounted for simply on the basis of behavioral accessibility, even though the scores of both groups on these subtests are highly related to it, because on each of these subtests the schizophrenics do significantly better at all levels of accessibility (Table 3). On Information and Block Designs the superiority of schizophrenic scores is less significant at the lowest level of accessibility than at the highest level. This indicates that, on these two subtests, as behavior becomes more inaccessible, the schizophrenics' performance more closely approaches that of the organics. Conversely, on the Thorndike-Binet Vocabulary the superiority of the schizophrenics' scores is inversely proportionate to accessibility. On this subtest the schizophrenics' scores are somewhat less vulnerable to poor accessibility than are those of the organics.

The significantly higher scores of the schizophrenics on these subtests cannot be attributed to differences in education, since the critical ratio for years of education in these two groups is only 0.56. The CR for length of hospitalization is also insignificant. However, a CR of 2.86 exists between the mean ages of these two groups, indicating that the organics are significantly older than the schizophrenics. While this factor must be considered, there is some evidence that age alone would not account for the difference, especially in all three subtests [26].

The conclusion, then, seems to be that Information, Block Designs, and Thorndike-Binet Vocabulary reflect thought processes that are impaired in direct proportion to behavioral in-

[3] In deriving these t's, the within-class variance was used as the error term.

---

because of the significant interaction variance and because of the shape of the curves [25], t tests were calculated. Those found to be significant are included in Table 3.

terference but that are more severely impaired by actual brain damage.

It is interesting that the Wechsler vocabulary completely fails to differentiate the two diagnostic groups while the Thorndike-Binet vocabulary does so at all levels of accessibility. Derivation of VQ's from the Wechsler-Bellevue weighted vocabulary scores shows that the Binet vocabulary is more difficult for organics than for schizophrenics. The schizophrenics' Wechsler VQ's are comparable to their Binet VQ's. However, the organics' Wechsler VQ's are significantly higher than their Binet VQ's. From inspection of the two vocabularies it appears that many words in the Binet vocabulary require greater elaboration and encompass more abstract concepts than do the Wechsler words. The organic patient's felt impotence and his excessive difficulty in conveying his thought in accurate verbalization [19] could account for the lower score on the Binet vocabulary.

On both Word Pairs I and Word Pairs II the schizophrenics do significantly better than the organics only at the upper level of behavioral accessibility. This suggests that organics show intrinsic difficulty in the thought processes of formation of, and especially in the retention of, new rote associations in addition to the deficit imposed by inaccessibility (Fig. 2). This also indicates that these processes are greatly disrupted by the emotional interference of schizophrenia.

On Digits Forward, Memory-for-Designs, Comprehension, and Sorting test, at the upper range of accessibility, the organics function at a level that is not significantly different from that of the schizophrenics. As behavioral accessibility decreases, the scores of both groups tend to show greater deficit (Fig. 1 and 2), but in general the diagnostic groups are not clearly differentiated. However, on Memory-for-Designs, the scores of the schizophrenics are significantly higher than those of the organics at the lower levels of accessibility. Thus, behavorial disturbance has greater disruptive effect on the thought processes involved in this subtest in the case of the organics than in the case of the schizophrenics.

The results on the Cube test indicate that a significantly greater percentage of organics than schizophrenics were able to benefit from the aids offered by that test and thereby to learn to construct the block designs, which could not be done without such aids. In view of this it would seem that the organic patient's difficulty on Block Designs is not primarily due to disturbance in analytic-synthetic thinking per se but to disturbance in attitude or shift which would be a component of such thinking.[4] Clinical observation reveals that without the aids of the Cube test the organic becomes fixated on one element of the design and often perseverates that operation in attempting to deal with other elements. Hence, he cannot transcend by himself the immediate situation and effect relationships which lie outside of it. The Cube test aids assist him in shifting his attitude so that he can learn to complete the tasks independently. On the other hand, it appears that the schizophrenic's emotional interference prevents him from goal-directed behavior and therefore from recognizing the purpose of the Cube test aids and from utilizing them.

Summary and Conclusions

Two test batteries were each administered to 42 organics and 85 schizophrenics on two successive days. No patient was of less than dull-normal intelligence or over 60 years of age. The test data were treated by dividing the patients into thirds according to their TRS scores and then analyzing the test scores according to variance attributable to diagnosis, to level of behavioral accessibility, and to the interaction of these components.

Within the limits of this study, the following conclusions seem tenable:

1. The test scores of psychotics primarily reflect behavioral inaccessibility rather than intrinsic deficit in the intellectual processes designated to be measured. Organic as well as schizophrenic disorganization of thought process is largely due to behavioral disturbance.

2. Deficits shown on Wechsler-Bellevue Vocabulary, Digits Reversed, Recall I, and Recall II reflect behavioral disturbance and not specifically brain damage. Application of this suggests that when used without qualification, many "deterioration" tests may measure behavioral rather than organic involvement.

3. Information, Block Designs, and Thorndike-Binet Vocabulary reflect thought processes that are impaired in direct proportion to behavioral interference but that are more severely impaired, at all levels of accessibility, in instances of actual brain damage.

4. Organics respond to the aids of the Cube test because these aids assist in shift of attitude, allowing the organics to learn to complete the task. Schizophrenics do not respond as well to these aids because of greater behavioral disturbance in this area.

Received January 2, 1953.

---

[4] Attitude and shift are discussed by Goldstein and Scheerer [6] and by Buerger-Prinz and Kaila in Rapaport [19].

References

1. Allen, R. M. The test performance of the brain injured. *J. clin. Psychol.*, 1947, **3**, 225-230.
2. Allen, R. M. The test performance of the brain diseased. *J. clin. Psychol.*, 1948, 4, 281-4.
3. Bleuler, E. *Dementia praecox or the group of schizophrenias.* New York: International Universities Press, 1950.
4. Boyd, F. A provisional quantitative scoring with preliminary norms for the Goldstein-Scheerer Cube Test. *J. clin. Psychol.*, 1949, 5, 146-153.
5. Goldstein, K. The significance of special mental tests for diagnosis and prognosis in schizophrenia. *Amer. J. Psychiat.*, 1939, **96**, 575-584.
6. Goldstein, K., & Scheerer, M. Abstract and concrete behavior: An experimental study with special tests. *Psychol. Monogr.*, 1941, **53**, No. 2 (Whole No. 239).
7. Graham, Frances K., & Kendall, Barbara S. Note on the scoring of the Memory-for-Designs test. *J. abnorm. soc. Psychol.*, 1947, **43**, 253.
9. Hanfmann, Eugenia, & Kasanin, J. Conceptual thinking in schizophrenia. *Nerv. ment. dis. Monogr.*, 1942, No. 67.
10. Hunt, H. F. *The Hunt-Minnesota test for organic brain damage.* Minneapolis: Univer. of Minn. Press, 1943.
11. Hunt, J. McV. Psychological loss in paretics and schizophrenics. *Amer. J. Psychol.*, 1935, **47**, 458-463.
12. Kasanin, J. S. (Ed.) *Language and thought in schizophrenia.* Los Angeles: Univer. of Calif. Press, 1944.
13. Lidz, T., Gay, J. R., & Tietze, C. Intelligence in cerebral deficit states and schizophrenia measured by Kohs Block Test. *Arch. Neurol. Psychiat.*, 1942, **48**, 568-582.
14. Lindquist, E. F. *Statistical analysis in educational research.* Boston: Houghton Mifflin, 1940.
15. Malamud, W., & Palmer, E. M. Intellectual deterioration in the psychoses. *Arch. Neurol. Psychiat.*, 1938, **39**, 68–81.
16. Mettler, F. A. (Ed.) *Selective partial ablation of the frontal cortex.* New York: Hoeber, 1949.
17. Rapaport, D. *Diagnostic psychological testing.* Vol. 1. Chicago: Year Book Publishers, 1945.
18. Rapaport, D. *Emotions and memory.* New York: International Universities Press, 1950.
19. Rapaport, D. *Organization and pathology of thought.* New York: Columbia Univer. Press, 1951.
20. Rappaport, S. R. The role of behavioral accessibility in intellectual function of psychotics. *J. clin. Psychol.*, 1951, **7**, 335-40.
21. Rashkis, H. A., Cushman, J. F., & Landis, C. A new method for studying disorders of conceptual thinking. *J. abnorm. soc. Psychol.*, 1946, **41**, 70-74.
22. Reichard, S., & Rapaport, D. The role of testing concept formation in clinical psychological work. *Bull. Menninger Clin.* 1943, **7**, 99-105.
23. Terman, L. M., & Merrill, Maud A. *Measuring intelligence.* Boston: Houghton Mifflin, 1937.
24. Thorndike, R. L. Two screening tests of verbal intelligence. *J. appl. Psychol.*, 1942, **26**, 128–135.
25. Webb, W. B., & Lemmon, V. W. A qualification in the use of analysis of variance. *Psychol. Bull.*, 1950, **47**, 130-137.
26. Wechsler, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.
27. Yacorzynski, G. An evaluation of the postulates underlying the Babcock Deterioration Test. *Psychol. Rev.*, 1941, **48**, 261-267.

Journal of Consulting Psychology
Vol. 17, No. 5, 1953

Use of The Performance Scale of the Wechsler Intelligence Scale for Children with the Deaf Child

E. Ellis Graham and Esther Shapiro
University of Denver

Most testing services are not prepared to test the deaf because of the scarcity of equipment, rarity of occasion, and expense involved. The purpose of this study was to investigate the possibility of using the Performance Scale of the Wechsler Intelligence Scale For Children (WISC) to obtain a valid measure of intelligence in deaf children. For this purpose, 20 deaf children were compared with two groups of hearing children, 20 in each group.

Much of the emphasis in the education of the deaf has been in preparing them vocationally so that they may be self-supporting. Concomitant goals are related to the achievement of more adequate adjustment: to help them enjoy a feeling of usefulness, to find group acceptance, to experience success as individuals. Although it is still controversial whether the auditorily handicapped are educationally retarded [4, 5, 6], the fact that the deaf child will eventually have to fit into the hearing child's world makes it imperative that we assess his abilities by the same standards that are used for the group with which he will compete. Since in most instances the deaf must compete with the hearing, it seems reasonable that tests for both groups be the same, modified only when necessary to exclude items dependent upon hearing.

The WISC, introduced in 1949, includes a Performance Scale which seems to lend itself to the modifications necessary for this type of investigation. It requires a minimum of verbalization, making it useful for nonhearing children who have not had the benefit of training in lip reading and it is becoming increasingly more acceptable in schools and clinics for use with the normal-hearing child; hence it is economical in that it is part of the standard equipment. Its measurement of various cognitive areas of intelligence for the normal-hearing child can provide a comparison with the deaf in whom it may be presumed to measure the same abilities. For these reasons the authors were interested in investigating the possibility of using this test with deaf children.

Procedure

In order to test the usefulness of the Performance Scale of the WISC with deaf children, three groups were set up. The members of Groups II and III, children with normal hearing, were individually matched according to certain pre-established criteria with the members of Group I, children with marked hearing disability.

Criteria for matching included hearing, physical health, sex, color and nativity, age, and intelligence. Normal hearing was defined simply as the absence of a known auditory defect; deafness was defined as a hearing loss of 60 db or greater, in both ears, sustained prior to the time of significant language developement. Good health of all children was established by health records, regularity of school attendance, and observation for apparent handicaps. The 9 girls and 11 boys were American-born white children, whose ages ranged between six years three months and twelve years two months. Subjects were matched according to age within three-month periods.

The Goodenough Draw-A-Man Test (DAM) was used as the basis for matching individuals with respect to IQ. This test was selected primarily because of its simplicity, both in content and in administration, its quick scorability, and because it does not make use of material included in the WISC. For the deaf group, the teacher or examiner requested that the child draw a picture of a man, the best picture that he could. As some of the younger deaf children did not understand these directions, the word MAN was printed on the blackboard in an enclosed frame, thus indicating to the child that he was to draw a picture. Standard directions [1] were given to the normal-hearing groups. The tests were scored and checked by two competent testers. In the case of disagreement, the papers were discarded. The IQ range was from 74 to 122 for the total number of subjects. Matching on the DAM was done within plus or minus five points, the probable error, as determined by Goodenough for her test [1].

The WISC Performance Scale was administered individually to the members of Groups I and II through pantomime instructions. Each child in the nonhearing group was introduced to the examiner by the teacher and told that he was going to spend some time with the examiner and play games. The hearing children were similarly introduced and told that the examiner wished to talk with them and play some games. The examiner instructed them that part of the game was that they must not talk and that if they watched closely, they would be able to understand what they should do. It was the examiner's impression that the children considered it a privilege to be called for testing and that rapport in all cases was good.

No difficulty was encountered in the administration of the Block Design, Object Assembly, or Coding tests since these appeared rather evident once the material was presented to the subject. The Picture Completion test seemed to present the most difficulty. After making several gestures to indicate that a part of the picture was missing, the examiner pointed out the missing parts in the first two pictures; this procedure was repeated until it was felt that the subject had grasped what he was to do. The initial four sets of cards in Picture Arrangement (designed for subjects under eight years of age) and the demonstration set were presented to Groups I and II. The demonstration cards were repeated with appropriate guestures until the subjects showed that they understood. Scoring for the Maze test was modified for Groups I and II, since the examiner could find no way of preventing the children from lifting their pencils from the paper. Accordingly, this misdemeanor was not penalized.

The subjects in Group III were introduced to the examiner either by the examiner or the teacher, and were given the WISC in accordance with the standardized instructions. The sixty WISC's were scored by the examiner in strict compliance with the scoring techniques [7], except for the modifications mentioned.

Table 1
Mean Scaled Scores

Test	Group I	Group II	Group III
*Total WISC (Performance)	96.1	95.5	101.15
Picture Completion	9.5	9.35	9.3
Picture Arrangement	8.8	9.5	11.6
Object Assembly	10.75	9.9	9.7
Block Design	9.85	9.95	9.45
Coding	8.7	8.5	10.75
Mazes	8.9	8.85	10.2

* Total here reported as IQ rather than scaled score.

Results and Discussion

Inspection of Table 1 shows that no consistent difference occurs in the mean scores of the various subtests in favor of any one group. When Group II is compared with Group III, statistically significant differences in favor of Group III are found for total Performance Scale and three of its subtests: Picture Arrangement, Coding, and Mazes. Between Groups I and III the total Performance Scale

Table 2
Significance of Differences between Groups (*t*)

Test	Groups I and II *t*	Groups I and III *t*	Groups II and III *t*
Total WISC (Performance)	.22	1.7	2.0*
Picture Completion	.15	.22	.05
Picture Arrangement	.91	2.8**	2.4*
Object Assembly	1.1	1.3	.46
Block Design	.14	.46	.66
Coding	.31	2.7*	2.7*
Mazes	.08	2.0*	1.4

* Significant at 5% level.
** Significant at 1% level.

is significantly higher for Group III, as are the scores for Picture Arrangement, Coding, and Mazes. Group I is significantly higher in performance on Object Assembly than is Group III.

Table 2 indicates that the differences between the performances of Groups I and II are not statistically significant. It will be noted that pantomime instructions were used in both these groups in an attempt to standardize the method of administration. However, the normal-hearing child may be handicapped in that he is not accustomed to expressing himself in pantomime or to interpreting pantomime directions given by others. This method, therefore, creates an artificial situation for him. For the deaf child, on the other hand, pantomime is a more natural means of expression; he has learned through the years to make himself understood by gestures and to understand others by watching for cues in facial and bodily expressions [2]. Group III was included to control the artifact which the pantomime instructions presented for the hearing child, and because the WISC was standardized with oral directions. Knowing that the three groups were equated in intelligence according to the Goodenough DAM, we cannot assume that Group III was brighter than either of the other groups. Consequently, it would seem that the difference in performance resulted from the method of administration. It is possible that the time factor, which operated in five of the six subtests, may have been influential in accounting for the differences between the groups receiving pantomime and verbal directions. Hiskey [3] believes it is unfair to include speed tests in a scale for deaf children as the concept of speed is difficult for the young deaf child to grasp. In general, this was found to be true, although in several instances the subjects seemed to understand.

Since pantomime seems inseperable from deafness, its effect must be considered in any test to be used for the deaf. Consequently, although the Performance Scale of the WISC as discussed in this study cannot be used validly to obtain a measure of intelligence for deaf children, it seems feasible that a correction factor which will nullify the effects of the pantomime instructions could be obtained. If such a factor can be established, then the test should be doubly valuable in assessing the deaf child's intellectual functioning. As it is, the test can be administered via pantomime as a crude measure.

Received January 15, 1953.

References

1. Goodenough, Florence L. *Measurement of intelligence by drawing.* Yonkers, N. Y.: World Book Co., 1926.
2. Heider, F., & Heider, Grace M. Studies in the psychology of the deaf. *Psychol. Monogr.,* 1941, 53, No. 5 (Whole No. 242).
3. Hiskey, M. S. A new performance test for young deaf children. *Educ. psychol. Measmt,* 1941, 1, 217-232.
4. MacKane, K. *A comparison of the intelligence of deaf and hearing children.* Teach. Coll. Contr. Educ. No. 585. New York: Teachers Coll., Columbia Univer., Bureau of Publications, 1933.
5. Myklebust, H. R. Clinical psychology and children with impaired hearing. *Volta Rev.,* 1948, 50, 55-56.
6. Stanton, Mildred B. *Mechanical ability of deaf children.* Teach. Coll. Contr. Educ. No. 751. New York: Teachers Coll., Columbia Univer., Bureau of Publications, 1938.
7. Wechsler, D. *Wechsler intelligence scale for children.* New York: Psychological Corp., 1949.

NEW BOOKS AND TESTS

Books

Barker, Roger G., Wright, Beatrice A., Meyerson, Lee, & Gonick, Mollie R. *Adjustment to physical handicap and illness: a survey of the social psychology of physique and disability.* (Rev. Ed.) Soc. Sci. Res. Coun. Bull. No. 55, Rev. 1954. Pp. xvi + 440. (paper) $2.00.

The effect of physical handicaps, cosmetic and functional, on attitudes and personality is well known, although generally little understood. The foreword to the first edition of this important survey quotes an illuminating passage from Sir Francis Bacon on the topic written over 300 years ago. The soma-psyche relations have recently had much elaboration in the various fields of psychosomatic medicine. This publication stresses the somatopsychological relations of physique and behavior with special reference to theory, research, and practice. Well-organized treatment is evident in the carefully outlined table of contents, extensive topical bibliography, careful author index, and detailed subject index. The volume is much too good to receive only casual comment. It should be read with studious care. The principle of adjustment *to* handicap rather than adjustment *of* the handicapped is clearly revealed as a significant reorientation which receives increasingly wide acceptance.—*E. A. D.*

Goldhamer, Herbert, & Marshall, Andrew W. *Psychosis and civilization: two studies in the frequency of mental disease.* Glencoe, Ill.: The Free Press, 1953. Pp. 126. $4.00.

This research monograph offers the best answer which the reviewer has yet seen to the often-asked question, "Is mental disorder on the increase?" The first of the two studies reported is based on a critical historical and statistical analysis of data covering a hundred-year period. It answers the question convincingly, "No." The second study, less cheerful in its conclusion, estimates that a person today has 1 chance in 20 of being affected by serious mental disorder by the age of 45, and 1 chance in 10 by the age of 65. The entire monograph deserves the careful attention of psychologists.—*L. F. S.*

Karpf, Fay B. *The psychology and psychotherapy of Otto Rank.* New York: Philosophical Library, 1953. Pp. xii + 129. $3.00.

This is a brief summary and sympathetic critique of Rank's theoretical contributions to psychoanalysis and psychology. The treatment is integrative and interpretative rather than narrowly factual, and the book makes pleasant and thought-provoking reading. —*W. A. H.*

Kessler, Henry H. *Rehabilitation of the physically handicapped.* (Rev. Ed.) New York: Columbia Univer. Press, 1953. Pp. xv + 275. $4.00.

The five years that have elapsed since the publication of the first edition of this book have shown us the continued need for the total utilization of manpower. There has been a steady growth in the recognition of the value and importance of rehabilitation. In the second edition Dr. Kessler has taken care to bring the illustrative material and tables up to date. There has been a complete rewriting of some of the chapters; for example, the one on vocational training, and the one on the blind and the deaf. The book should continue to occupy its important position as a source book for those working with the mentally and physically handicapped. There is an excellent index and a list of major centers and agencies for the handicapped.—*B. M. L.*

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953. Pp. xix + 393. $6.50.

Lindquist's most recent textbook on design is essentially a modernized, more sophisticated version of his widely-used *Statistical Analysis in Educational Research* (1940). Six main types of design are considered in detail, together with their variants and mixtures: simple-randomized, treatments by levels, treatments by subjects, random replications, factorial, and groups-within-treatments. The au-

---

Note.—The reviews were prepared by the Editor and the Associate Editors, who may be identified by their initials.

thor's plan, to begin the exposition of each design with basic principles and mathematical derivations instead of concrete illustrations, makes for hard reading but prevents the student from looking for mere recipes. Ample exercises follow each chapter. The book is best used as a text in conjunction with competent class instruction. It can also serve as a stimulating reference for psychologists with a good command of basic statistics.—*L. F. S.*

Mowrer, O. Hobart. (Ed.) *Psychotherapy: theory and research.* New York: Ronald Press, 1953. Pp. xviii + 700. $10.00.

Twenty-two psychologists collaborated in the preparation of this impressive volume which combines eight chapters of essays on methods and theory, 201 pages, with eleven chapters, 455 pages, on research techniques and findings. The chapters on theory vary greatly in their nature, but one cannot imply that they are uneven in quality; each is a good and typical example of its author's work. To refer to only a few: May gracefully relates therapy to its philosophical background; Rogers illustrates some of his convictions with excerpts from interviews; Mowrer presents a theory of neurosis and therapy synoptically. The most notable characteristic of the volume, in contrast with most other books on psychotherapy, is that it devotes more than two-thirds of its length to research studies. All of the materials are meaty. Some are summaries or extensions of research programs already reported, as in Seeman and Raskin's chapter on research in client-centered therapy, and in Dollard's and Mowrer's separate sections on measuring tension in written documents. There are methodological contributions by Mowrer, Cronbach, McQuitty, and others. Research results not reported elsewhere include studies of verbal behavior and of tension changes during psychotherapy, carried out by Mowrer and his students. The volume as a whole is hard to categorize. It is not exactly a text, a treatise, or a research report, but is something of all three. It is decidedly worth reading, by psychologists and by their advanced students.—*L. F. S.*

Overholser, Winfred. *The psychiatrist and the law.* New York: Harcourt, Brace, 1953. Pp. x + 147. $3.50.

In their roles both as citizens and as professionals, psychologists will find this little book stimulating and informative. Winner of the Isaac Ray Award for the most worthy "contribution to the improvement of the relations of Law and Psychiatry," it presents a provocative social psychological analysis of the law as an institution, trenchantly examines the legal issues involved in the commitment, institutional maintenance, and release of patients in psychiatric hospitals, and discusses with great insight the functions and problems of the student of behavior as an expert witness. Dr. Overholser writes out of rich personal experience, with a genuine scholar's large familiarity with legal precedents and the relevant literature in the history of both psychiatry and law, and from a broad and humane point of view. His contribution to interdisciplinary understanding is considerable, and his illustration of how one profession deeply concerned with human conduct can usefully and constructively criticize another provides a model to which the other behavioral sciences might well pay heed. It is regrettable that his keen thinking was not directed toward those legal matters bearing on interprofessional relationships, and one wonders if he does not protest too much in his repeated insistence that all legally relevant mental states constitute *medical* questions.—*E. J. S.*

Podolsky, Edward. (Ed.) *Encyclopedia of aberrations.* New York: Philosophical Library, 1953. Pp. ix + 550. $10.00.

This volume is the result of the efforts of fifty-one physicians most of whom are psychiatrists, seven psychologists, and one chemist. The articles (or original papers, as Alexandra Adler calls them in a foreword) vary in length from a single line to many pages. Twenty-six pages, for example, are devoted to the psychopathic personality. The topics are mostly well-known symptoms and syndromes. There are also definitions of less familiar terms as well as discussions of a few basic concepts. A sample of topics includes aboiement, anti-Semitic attitudes, anxiety states, disorientation of body image, coarctation, folie à deux, gelasmus, nonadaptive group behavior, nuns' melancholy, orality, behavior during puberty, and sophomania. The organization within the topics varies greatly. Some topics include long cases, a few contain tables, diagrams, and Rorschach summaries. Bibliographies and references to other publications are extremely rare. Many articles deal with theoretical explanations of the phenomena. The writing style is quite readable; in fact, almost popular in nature. The reviewer misses an index or lengthy table of contents and feels the absence of it will reduce efficiency in the use of the book. There also seem to be few cross-references. Some of the topics represent an excellent review of the phenomena.—*F. McK.*

Reid, Joseph H., & Hagan, Helen R. *Residential treatment of emotionally disturbed children.* New York: Child Welfare League of America, Inc., 1952. Pp. ix + 313. $3.50.

In the more severe behavior disturbances of children, and those where family life is not conducive to recovery at home, therapeutic programs of extrafamilial procedures have been often used. This study reports a descriptive survey of twelve residential treatment centers, in order to provide guidance in establishing similar programs elsewhere. Five of these centers represent social agency programs, and seven are essentially medical (psychiatric), representing the major developments in this field in the United States. Three are supported by public funds and nine by private and community funds, one primarily by patient fees. The study deliberately avoids evaluation, concentrating instead on such issues as organization, plant and grounds, intake policies and procedures, population, staff, treatment program, costs, personnel training. The report is factually informing as to resources and practices. Professional readers will sense numerous implications and will no doubt derive evaluations of their own. Psychologists will be specially interested in the opportunities for service afforded and the conditions of professional and interprofessional collaboration.—*E. A. D.*

Sullivan, Harry Stack. *The interpersonal theory of psychiatry.* New York: Norton, 1953. Pp. xviii + 393. $5.00.

Harry Stack Sullivan's approach to psychiatry, as expressed in *The Interpersonal Theory of Psychiatry,* will probably be more acceptable to psychologists than that of any other well-known psychiatrist. He arrived at his destination via clinical research in schizophrenia and analytic therapy in the psychoneuroses; the vehicles were psychoanalysis, Adolph Meyer's psychobiology, field theory, and social psychology. The result is an eclecticism that is dynamic and significant. Sullivan rejects the concept of psychiatry as merely the study and treatment of abnormal behavior; much more important, as far as he is concerned, is the study of interpersonal relationships. This work is an edited version of Sullivan's lectures and notebooks from 1946 to 1949, the year of his death. Since his lectures were more sparkling than his writing, this volume, frequently in colloquial vein, makes interesting and relatively easy reading.—*M. K.*

Whiting, John W. M., & Child, Irvin L. *Child training and personality: a cross-cultural study.* New Haven: Yale Univer. Press, 1953. Pp. vii + 353. $5.00.

An important research report on the interrelations between personality and culture as seen in data about 75 societies, obtained mainly from the Human Relations Area Files of the Institute of Human Relations. The orientation is toward the testing of general hypotheses, drawn from psychoanalysis and general behavior theory, rather than toward the understanding of any one society. A correlational method is used, each culture being considered as a unit. The results, which should be examined in detail, give little support to some hypotheses and much to others. In general, they confirm the need to divide broad psychoanalytic formulations into more detailed hypotheses restated in terms of general behavior theory. No one should teach a course or write a book on personality without considering the evidence preesnted here.—*L. F. S.*

Tests

Bennett, George K., & Gelink, Marjorie. *Short Employment Tests, 1953 Manual.* New York: Psychological Corp., 1953. Pp. 10. 35¢.

The battery of three five-minute tests for clerical applicants—verbal, numerical, and clerical aptitude—was issued in 1951 with a preliminary manual (see *J. consult. Psychol.,* 1952, **16,** 159). The new manual contains evidence of the validities of the tests in terms of various relevant criteria, obtained from six different businesses. There are also tables of reliabilities, intercorrelations, relations to other tests, and norms.—*L. F. S.*

Bruce, Martin M. *Sales Comprehension Test, Form M.* Adults. 1 form. Untimed, (20) min. Test booklet (20¢), with manual, pp. 8; specimen set ($1.00).

Business Judgement Test. Adults. 1 form. Untimed, (15) min. Test booklet (15¢), with manual, pp. 4; specimen set (75¢).

Sales Motivation Inventory. Adults. 1 form. Untimed, (30) min. Test booklet (15¢), with manual, pp. 4; specimen set (75¢). New York 9 (624 E. 20 St.): Author, 1953.

These three business tests are of special interest because of the author's explicit plan to report data according to the recommendations of the APA Committee on Test Standards. As a result, the manuals are more informative than are many such documents, and candidly show the limitations as well as the values of the tests. *The Sales Comprehension Test* is the best-developed of the three, being based on item analyses and other studies carried out over a period of several years. It discriminates salesmen and nonsalesmen successfully, and has some relationship to success in selling. The other two tests are not as fully developed, but their manuals show some preliminary evidence of usefulness, and are free from undue claims.—*L. F. S.*

Wood, Hugh B. *Behavior Preference Record.* 3 levels, Gr. 4-6, 7-9, 10-12. 2 forms at each level. Untimed, (30-45) min. IBM, Scoreze, or hand scoring. Test booklet (8¢); answer sheet (4¢); Scoreze (7¢); with manual, pp. 16; specimen set (35¢). Los Angeles: California Test Bureau, 1953.

This questionnaire presents 15 to 20 described interpupil situations in school settings, each followed with multiple-choice alternatives for responses to "what would you do?" and "because." It is scored for five variables which are called cooperation, friendliness, integrity leadership, and responsibility, and for a critical thinking score presented as an index of the degree of rationalization indicated in the "because" choices. Statistical data are given fully, including (for the Grade 10 to 12 forms), trait reliabilities ranging from .72 to .90, correlations with teacher's ratings ranging from .46 to .78, and trait intercorrelations from —.03 to + .34. Norms are based on "some seventeen hundred" cases, which are unspecified with respect to levels and forms. While the blank may evoke some useful thinking about the social characteristics of pupils, it suffers from numerous faults. The prestige values of the responses were not controlled. As in all such short questionnaires, radically different interpretations from "average" to "very high" may be drawn from scores which differ by as little as 3 points when the standard error of measurement is as great as 2.8 points.—*L. F. S.*

Books Received

Berkman, Tessie D. *Practice of social workers in psychiatric hospitals and clinics.* New York (1860 Broadway): Amer. Assoc. of Psychiatric Social Workers, 1953. Pp. ix + 158. (paper) $2.00.

Buros, Oscar Krisen. (Ed.) *The fourth mental measurements yearbook.* Highland Park, N. J. (220 Montgomery St.): Gryphon Press, 1953. Pp. xxiv + 1163. $18.00.

Hammond, Kenneth R., & Allen, Jeremiah M., Jr. *Writing clinical reports.* New York: Prentice-Hall, 1953. Pp. xii + 235. $5.35.

Heiman, Marcel. (Ed.) *Psychoanalysis and social work.* New York: International Universities Press, 1953. Pp. xiv + 346. $5.00.

Kubie, Susan H., & Landau, Gertrude. *Group work with the aged.* New York: International Universities Press, 1953. Pp. 214. $3.50.

Lacey, Oliver L. *Statistical methods in experimentation.* New York: Macmillan, 1953. Pp. xi + 249. $4.50.

Lindgren, Henry Clay. *The psychology of personal and social adjustment.* New York: American Book Co., 1953. Pp. ix + 481. $4.50.

Moreno, J. L. *Who shall survive?* Beacon, N. Y.: Beacon House, 1953. Pp. cxiv + 763. $10.00.

Naumburg, Margaret. *Psychoneurotic art.* New York: Grune & Stratton, 1953. Pp. x + 148. $6.75.

Peters, R. S. (Ed.) *Brett's history of psychology.* New York: Macmillan, 1953. Pp. 742. $7.50.

Riese, Walther. *The conception of disease.* New York: Philosophical Library, 1953. Pp iv + 120. $3.75.

Roheim, Geza. *The gates of the dream.* New York: International Universities Press, 1952. Pp. x + 554. $10.00.

Sabine, Paul E. *Atoms, men and God.* New York: Philosophical Library, 1953. Pp. x + 256. $3.75.

Townsend, John C. *Introduction to experimental method.* New York: McGraw-Hill, 1953. Pp. ix + 220. $4.00.

Westwood, Gordon. *Society and the homosexual.* New York: Dutton, 1953. Pp. 191. $3.00.

Woolf, Maurice D., & Woolf, Jeanne A. *The student personnel program.* New York: McGraw-Hill, 1953. Pp. ix + 416. $5.00.

Journal of Consulting Psychology
Vol. 17, No. 6, 1953

Ranking Bellevue Subtest Scores for Diagnostic Purposes[1]

Joseph Jastak
University of Delaware[2]

One of the features of the Bellevue scale is the diagnostic significance of rank patterns of the 11 subtests. Wechsler [4] has found that the pattern of sigma deviations of individual subtest scores from the mean is to some extent related to psychiatric disease entities. His schizophrenics, for example, rank high in information and vocabulary. They rank low in picture completion and similarities. His psychopaths rate high in object assembly and picture arrangement, low in information and arithmetic. Wechsler gives sign lists for neurotic, schizophrenic, organic, psychopathic, and mentally defective patients [4, pp. 150-152]. Clinical impressions substantially agree with the ability patterning reported by Wechsler. However, there is no objective way of applying his sign lists to individual case records. If an objective method were applied, the signs would not withstand the rigors of clinical and experimental testing.

This study was undertaken to find the possible reasons why the sign lists fail. We assume, along with Wechsler, that ability ranks may provide a sound basis for the study of personality differences [1]. We also assume that such differences are measurable, and that their measures may be clinically useful.

Problems and Methods

The measurement of subtest dispersion has two aspects: (*a*) the rank orders in which measured abilities occur and (*b*) the extent to which their scores deviate from a particular reference point. Our study is confined to the first of the two aspects. To measure rank orders and their patterning, we have found the rank-difference correlation method (rho) most convenient. The rho method has a number of advantages. It avoids dealing with plus and minus entries. Its coefficient directly represents the multiple relationships between subtests without intervening reference points like the mean or vocabulary scores. The ranks are independent of ability levels. Inferior and superior intellects may have similar or identical ranks. Ranking is also independent of the degree of discrepancies between scores which can be measured by other means. The rho makes the comparison of individual records with each other or with those of a group a relatively simple matter, and it facilitates cross validation. Its probable error is fairly well established.

Some years ago, we analyzed the Wechsler-Bellevue records of several population samplings by the rho method. Police and nurse applicants, state hospital employees, and neurotic and psychotic patients of all types were sampled in fair-sized groups. The obtained subtest ranks were all highly correlated. None of the rho's between group patterns was lower

---

[1] This study is a by-product of more extensive research on mental diagnosis undertaken with financial aid from the Kate Jackson Anthony Foundation of Lewiston, Maine. The Foundation's generous support is here gratefully acknowledged.

The numerous calculations were facilitated by the use of a Monroe automatic calculator which was made available to the author through the efforts of the Rev. H. E. Hammond of Old Swedes Church and the Christina Community Center, Wilmington, Delaware.

A modification of this paper was presented at the annual meeting of the Delaware Psychological Association on October 13, 1952.

[2] The psychometric data for this study were assembled and tabulated while the author was chief psychologist at the Delaware State Hospital and Mental Hygiene Clinic, Farnhurst, Delaware.

Table 1

Correlation Coefficients (rho's) of Subtest Ranks of Bellevue Weighted Score Means for Seven Diagnostic Groups Studied by Rapaport

Groups	*N*	1	2	3	4	5	6	7
1. Unclassified schizophrenics	37		.76	.76	.77	.66	.68	.68
2. Paranoid schizophrenics	26			.66	.87	.76	.76	.73
3. Paranoid condition and simple schizophrenics	22				.76	.46	.77	.61
4. Preschizophrenics	32					.78	.92	.83
5. Depressives	31						.70	.67
6. Neurotics	59							.84
7. Police patrols	54							

than .83. Ranking of subtests failed to differentiate between normal and abnormal groups or between any two abnormal groups.

Then we ranked the Bellevue records of studies published by others and got results very much like our own. The intercorrelations tended to be significantly positive with few exceptions. The individual coefficients varied more with size and homogeneity of the group than with the diagnosis. As an example, Table 1 contains the correlation matrix for the weighted score ranks of seven diagnostic groups published by Rapaport [3].

The high positive intercorrelations between Rapaport's groups support our findings concerning pattern similarity in highly different groups. When Rapaport's main groups are split into their subcomponents, only his simple schizophrenics ($N = 9$) and his maladjusted patrols ($N = 9$) deviate from the trend of high positive correlations. A comparison of Rapaport's seven group patterns with five of our own patterns gave 35 rho's ranging from .66 to .95, indicating that Bellevue subtest patterns tend to be similar irrespective of place, time, and population sampling. We consider the bunching of rho's at the high positive end unusual, since there is no logical reason why they should not vary through their entire possible range from minus one to plus one.

If a schizophrenic and a psychopathic patient have antithetic ability ranks, as is often reported and assumed to be the case, then the rho coefficient between their score ranks should be high and negative. This does not appear to be the case. In view of such puzzling results we have asked ourselves the following questions. Is the idea of differentiating ability ranks between diagnostic groups a false one? Are our methods of measuring ability patterns inadequate? Are external criteria of grouping unreliable? Should we look for factors obscuring meaningful ability patterns? Is there perhaps something in the norms of the Bellevue scale that makes ability ranks so uniform?

First, let us see to what an extent Wechsler's own sign lists are correlated. In Table 2 we present the matrix of rho's for the five diagnostic groups described in Wechsler's book.

Table 2

Intercorrelations (rho's) of Wechsler's Ranked Sign Lists for Five Diagnostic Groups*

Group	Schizophrenic	Neurotic	Psychopathic	Mental defective
Organic	.37	.63	.06	.33
Schizophrenic		.65	–.13	.13
Neurotic			–.14	.27
Psychopathic				.47

* [4, pp. 150-152]

In Table 2 we see that only three out of ten entries are significantly positive and that these are not as large as ours. Moreover, the psychopathic pattern is actually negatively related to the neurotic and schizophrenic patterns. Most of the rho's are not appreciably different from zero which is as it should be.

To find out why Rapaport's and our own rho's are so high, we have made five different analyses of the test ranks of five groups. We have used the Jastak-Bijou Wide Range Reading test [2] in addition to the Bellevue subtests. This gives us 12 subtests and 12 ranks for each group and each method of analysis.

The five groups were: (*a*) Delaware State Hospital employees (attendants, office clerks, matrons, occupational aides, technicians, maintenance and recreation workers), (*b*) patients diagnosed as psychoneurotic, (*c*) patients diagnosed as manic-depressive, (*d*) patients diagnosed as schizophrenic, and (*e*) patients diagnosed as psychotic because of organic brain disease. No alcoholics were included in the group of organic cases. The ranges, means, and standard deviations of the ages of our groups are shown in Table 3. There are 80 cases in each group, 40 males and 40 females, except in the organic group which has 40 males and 34 females, a total of 74 cases.

Table 3

Means, Standard Deviations, and Ranges of Ages in Diagnostic Groups

Diagnostic group	*N*	Range	M	*SD*
Nonpsychotic				
Male	40	18–68	35.02	13.42
Female	40	19–58	32.60	11.06
Neurotic				
Male	40	19–65	32.40	11.00
Female	40	20–64	36.78	12.39
Manic-depressive				
Male	40	18–67	38.82	13.69
Female	40	20–69	41.25	12.85
Schizophrenic				
Male	40	17–47	30.25	7.26
Female	40	20–68	33.10	9.89
Organic				
Male	40	19–77	43.32	15.65
Female	34	18–67	43.94	11.33
Total Number	394			

The results of our five groups were analyzed in the following ways:

1. *Weighted score method.* Bellevue weighted scores of each subtest were averaged for each group. The obtained averages were ranked according to size from highest (rank 1) to lowest (rank 12). The resulting rank orders were then correlated by the rho method. The use of weighted scores makes the results entirely dependent on Wechsler's normative universe without any known controls.

2. *Quotient method.* The weighted score of each subtest for every person was multiplied by five. A quotient corresponding to this product was found in Wechsler's IQ Tables for verbal or performances tests in the appropriate age column. The means of the quotients for each group were calculated, rânked, and correlated as before. This method depends on Wechsler's norms, but introduces age control since quotients tend to neutralize changes due to the age factor.

3. *Independent sampling method with no controls.* For this method we calculated the average raw scores on each subtest obtained by 1,241 individuals. Of these, 329 were state hospital employees, 169 psychoneurotic patients, 181 nonpsychotic patients, 101 patients of unknown diagnosis, and 461 psychotic patients of all kinds. The mean raw scores on each subtest of our five experimental groups were compared with the corresponding subtest means of the total group by sigma deviations. The latter were ranked in order from largest plus to largest minus deviation, or from largest plus to smallest plus deviation, or from smallest minus to largest minus deviation. The resulting ranks were correlated with each other. This method is independent of Wechsler's test norms.

4. *Independent sampling method with age control.* In this step we divided the 1,241 cases into four groups according to age: 440 cases from 17 to 29 years, 330 cases from 30 to 39 years, 273 cases from 40 to 51 years, and 198 cases from 52 to 77 years. We computed the mean raw scores for each subtest and age group. Then we found the differences between these means and those of our five groups separated by age, and divided them by the sigmas of the original age groups. The obtained ratios were averaged, ranked, and correlated with each other in each of the five groups. This method is independent of Wechsler's norms and, in addition, tends to control roughly the most important changes due to aging. It partly eliminates test differences between persons or groups of varying age levels and thereby emphasizes the residual relationships due to personality differences or psychiatric illness.

5. *Independent sampling method with age and sex control.* We next divided each of the four age groups under 4 into two parts according to sex. The mean raw scores for each subtest in the resulting eight groups served as norms from which sigma deviations were calculated. These were ranked as before and cor-

related by the rho method. This analysis should neutralize any sex differences which might exist in ability ranks.

Results

In reporting results, male and female groups will be treated separately through all steps. This will permit us to evolve male and female patterns if such should be justified. Also, if in the course of the analysis, the observed changes are parallel for both sexes, our conclusions concerning the conditions under which pattern analysis is most effective will have greater weight than they would have if observed in one group only. The results of the weighted score method are presented in Table 4.

Table 4
Correlation Coefficients (rho's) of Subtest Ranks of Bellevue Weighted Scores for Five Diagnostic Groups*

Group	Neurotic	Manic-depressive	Schizophrenic	Organic
Male				
Employees	.77	.62	.61	.29
Neurotic		.80	.89	.65
Manic-depressive			.92	.76
Schizophrenic				.81
Female				
Employees	.12	–.18	.14	.07
Neurotic		.82	.83	.88
Manic-depressive			.89	.91
Schizophrenic				.85

* For males, N=200, 40 in each group; for females, N=194, 40 in each group except for 34 in organic group.

The matrix of Table 4 has significantly positive rho's in 15 out of 20 cells. The ability ranks of female employees are apparently different from those of the other female groups. The patterns of male employees and organic patients are also different. We will consider rho's of .40 and below not significantly different from zero. The coefficients of neurotic and psychotic groups for both sexes are so highly positive that their test ranks cannot be clearly distinguished from each other.

The rho's of the quotient method are given in Table 5. The values of Table 5 are generally somewhat smaller than are those of Table 4.

Table 5
Correlation Coefficients (rho's) of Subtest Ranks of Bellevue Quotients for Five Diagnostic Groups

Group	Neurotic	Manic-depressive	Schizophrenic	Organic
Male				
Employees	.72	.56	.34	.07
Neurotic		.76	.77	.59
Manic-depressive			.85	.61
Schizophrenic				.83
Female				
Employees	.25	–.15	–.01	.13
Neurotic		.74	.73	.74
Manic-depressive			.94	.74
Schizophrenic				.77

However, they are still significantly positive in 14 of 20 cells. The relationship between male employees and schizophrenics has declined sufficiently to permit the separation of the two patterns. In other respects, the quotient method yields similar rank orders for neurotics and psychotics of both sexes.

The results of Table 6 were obtained from our third analysis, the independent sampling method with no age or sex control. The coefficients of Table 6 are, with one exception, considerably lower than those of Table 4 or 5. The relationships between employee and patient ranks have become negative in six out of

Table 6
Correlation Coefficients (rho's) of Subtest Ranks of Bellevue Raw Scores for Five Diagnostic Groups

Group	Neurotic	Manic-depressive	Schizophrenic	Organic
Male				
Employees	.44	–.26	–.49	–.48
Neurotic		.09	.24	.12
Manic-depressive			.77	.56
Schizophrenic				.71
Female				
Employees	.21	–.45	–.23	–.31
Neurotic		.54	.34	.21
Manic-depressive			.78	.67
Schizophrenic				.88

eight cells. Male neurotic patterns can now be clearly distinguished from those of manic-depressives, schizophrenics, and organics. Female neurotics are distinguishable from female schizophrenics and organics. Twelve out of 20 entries represent different patterns (rho's below .40).

The results from our fourth analysis with age control are listed in Table 7. Age control has left most intercorrelations unchanged in comparison with those of Table 6. The most striking change is the reduction of rho's between manic-depressives and organics in both male and female groups. The number of distinct diagnostic ranks has been raised from 12 to 14 as a result of age control.

Table 7

Correlation Coefficients (rho's) of Subtest Ranks of Bellevue Raw Scores for Five Diagnostic Groups with Age Control

Group	Neurotic	Manic-depressive	Schizophrenic	Organic
Male				
Employees	.29	–.11	–.41	–.39
Neurotic		.30	.38	.09
Manic-depressive			.68	.17
Schizophrenic				.62
Female				
Employees	.21	–.24	–.24	–.05
Neurotic		.51	.36	.28
Manic-depressive			.82	.50
Schizophrenic				.76

Table 8

Correlation Coefficients (rho's) of Subtest Ranks of Bellevue Raw Scores for Five Diagnostic Groups with Age and Sex Control

Group	Neurotic	Manic-depressive	Schizophrenic	Organic
Male				
Employees	.11	–.45	–.59	–.68
Neurotic		.03	.31	–.43
Manic-depressive			.42	.08
Schizophrenic				.57
Female				
Employees	–.07	–.83	–.78	–.50
Neurotic		.20	–.13	–.06
Manic-depressive			.66	.32
Schizophrenic				.48

The rank correlations from our independent sampling with age and sex control are recorded in Table 8. Sex control has increased the number of distinct rank orders from 14 to 16. Only four cells out of 20 have high positive correlations. Two of these are of borderline significance. We can now distinguish well between all diagnostic groups except manic-depressives and schizophrenics, schizophrenics and organics in both sex groups. The high negative correlations between employees and all psychotic groups are noteworthy.

We investigated the possible reasons for the consistently positive correlations between schizophrenics and manic-depressives and between schizophrenics and organics. Two explanations seem plausible. We have found that a large number of incipient schizophrenics are diagnosed as manic-depressive. When we separated the younger groups of schizophrenics from the older patients of the same category and correlated the resultant two patterns with that of manic-depressives, we found that the correlation between old schizophrenics and manic-depressives became negative (—.24) and that the correlation between young schizophrenics and manic-depressives became more positive (.76). Psychiatrists consulted on this point agreed that the high relationship was not due to the similarity between young schizophrenics and manic-depressives but due to inaccurate diagnoses. A number of those diagnosed as manic or depressive turned out to be clear-cut schizophrenics. A possible explanation of the positive relationship between schizophrenics and organics is that many of the schizophrenic patients were examined at varying periods following electric shock therapy which is known to change temporarily the original functional pattern to an organic one.

In Table 9 we list the ranks of the Bellevue subtests and the WRAT Reading test for our five groups according to sex. These ranks were obtained from the fifth analytic step, the rho's of which appear in Table 8. These ranks can*not* be obtained by using deviations expressed in terms of Wechsler's weighted scores or quotients. The data in Table 9 reveal that the

Table 9

Raw Score Ranks of Wechsler-Bellevue Subtests and Wide Range Reading Test for Five Diagnostic Groups Obtained from Independent Sampling of 1,241 Records, with Age and Sex Controlled ($N=40$ in each group except for 34 female organics)

Test	Employees		Neurotic		Manic-Depressive		Schizophrenic		Organic	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
1. Information	11	8	6	5	4	6	1	3	3	3
2. Comprehension	3	7	3	1	8	8	6	10	9	6
3. Digit Span	8	10	12	11	12	4	7	4	2	8.5
4. Arithmetic	7	9	11	2	2	5	8	7	5	8.5
5. Similarities	2	5	5	3	6	3	5	6	7	11
6. Vocabulary	12	11	1	4	1	2	3	2	8	2
7. Picture Arrangement	9	1	8	7	9	12	10	9	4	4
8. Picture Completion	6	3	2	8	11	7	4	12	6	7
9. Block Design	5	4	7	10	5	11	9	5	11	10
10. Object Assembly	4	6	10	12	7	9	11	8	10	5
11. Digit Symbol	1	2	4	6	10	10	12	11	12	12
12. WRAT Reading	10	12	9	9	3	1	2	1	1	1

WRAT Reading test holds a low rank in non-psychotics and gradually rises to first rank with increasing disorganization or severity of the mental condition. A similar progression from low to high rank occurs in the information and vocabulary tests, though neither of them is as clear-cut as that of the reading test. The digit-symbol, block design, and similarities tests tend to decline in rank with increase in mental disturbance. The digit-symbol test is as definite in its decline as the reading test is in its ascent. The remaining six subtests do not show consistent rises or declines but vary with the diagnostic group.

The rank patterns of the two sexes within each group are positively correlated with each other. There are, however, some striking sex differences in each diagnostic group in at least one subtest. These differences are, in our view, quite consistent and may well disturb diagnostic patterns derived from a sexually mixed universe.

In order to enable psychologists to apply our findings to individual clinical cases, we reproduce in Tables 10 and 11 the means and standard deviations of the raw scores of each subtest for ten groups separated by age and sex. Only eight of the ten groups (from 17 years up) were used in our analysis, but the data for boys and girls between the ages of 13 and 16 are added for the benefit of those who might wish to use them. The patterns of each diagnostic group are highly correlated at different ages. Thus the rank orders of neurotics between the ages of 20 and 25 are similar to the rank orders of neurotics between the ages of 50 and 59.

The facts of Table 10 or 11 may be used in the following manner. 1. Administer the Wechsler-Bellevue scale and the Jastak-Bijou Wide Range Reading test in full to get a raw score for each of the 12 subtests. 2. Subtract the mean of the subtest in Table 10 or 11 from the obtained raw score on that subtest, being careful to enter the column appropriate to the age and sex of the case. If the patient's raw score is larger than the mean, the difference will be positive. If the patient's raw score is smaller than the group mean, the difference will be negative. 3. Divide the resultant difference for a subtest by the standard deviation in Table 10 or 11 for that subtest. 4. Rank the 12 deviation ratios from largest positive to smallest positive, or largest positive to largest negative, or smallest negative to largest negative, as the case may require. 5. Correlate the patient's rank pattern with each of the five patterns printed in Table 9, using the ranks corresponding to the sex of your client.

Though our groups are small and lack cross validation, the largest positive coefficient will frequently place the patient in the appropriate diagnostic group. At times, all five coefficients will be so close to zero that none will fit the reported patterns. The following cautions should be observed in the interpretation. A

Table 10
Means and Standard Deviations of Subtest Raw Scores on the Wechsler-Bellevue Scale and on the Jastak-Bijou Wide Range Reading Test for Patients of Ages 13 through 39

	Ages 13–16 yrs.				Ages 17–29 yrs.				Ages 30–39 yrs.			
	Male (*N*=110)		Female (*N*=102)		Male (*N*=275)		Female (*N*=165)		Male (*N*=192)		Female (*N*=138)	
Test	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1. Information	8.40	4.21	6.97	4.28	13.06	4.89	10.58	4.52	13.04	5.28	10.83	5.09
2. Comprehension	7.56	2.91	7.08	3.40	9.99	3.58	8.59	3.43	10.25	3.71	8.54	3.51
3. Digit Span	9.32	1.88	9.17	2.02	10.15	2.47	10.12	2.11	10.03	2.21	9.83	2.26
4. Arithmetic	4.87	2.31	4.16	2.29	6.84	2.95	5.45	2.46	7.02	2.76	5.33	2.48
5. Similarities	7.06	3.33	7.35	4.78	9.81	4.53	9.42	3.81	9.34	4.74	9.70	4.77
6. Vocabulary	13.46	4.84	12.91	5.50	19.84	7.23	18.72	6.14	21.11	7.93	20.99	7.49
7. Picture Arrangement	8.35	3.69	8.35	3.98	9.62	4.09	9.04	4.22	8.05	3.73	7.28	4.11
8. Picture Completion	8.96	2.41	7.44	2.81	9.92	2.85	8.29	3.02	9.60	2.89	7.97	3.16
9. Block Design	17.24	7.02	14.10	8.15	19.84	8.57	17.07	8.38	17.43	7.42	15.04	8.31
10. Object Assembly	16.74	3.86	15.17	4.94	17.17	4.40	16.09	4.52	15.70	4.78	15.72	4.74
11. Digit Symbol	30.98	8.84	35.60	10.83	35.26	12.06	36.13	12.26	30.02	12.53	30.36	14.40
12. WRAT Reading	52.37	20.89	56.91	19.66	74.39	23.04	75.80	19.10	76.58	23.34	76.48	26.46

high correlation with the nonpsychotic order does not necessarily mean normalcy. Our rank patterns do not take into consideration the degree of displacements between test scores which are very important in judging the disturbance of the individual. Electric shock and other treatments may radically change ability patterns. The psychiatric diagnosis has a large margin of error, especially at the younger age levels. There are many other ability patterns which are not included in this study. Each major group, like the neurotic, appears to have two or more subpatterns which are diagnostically important. The patterns we list may therefore be combinations of several subpatterns.

Summary and Conclusions

We are now ready to answer the questions asked in the beginning of this paper. Our results generally confirm the correctness of the assumption that ability patterns and psychiatric

Table 11
Means and Standard Deviations of Subtest Raw Scores on the Wechsler-Bellevue Scale and on the Jastak-Bijou Wide Range Reading Test for Patients of Ages 40 through 82

	Ages 40 – 51 yrs.				Ages 52+ yrs.			
	Male (*N*=171)		Female (*N*=102)		Male (*N*=114)		Female (*N*=84)	
Test	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1. Information	12.32	5.26	10.42	5.03	11.04	5.07	9.33	4.78
2. Comprehension	9.79	3.66	8.28	3.51	8.94	3.76	7.23	3.76
3. Digit Span	9.80	2.19	9.68	2.34	9.35	2.22	9.39	2.28
4. Arithmetic	6.67	2.87	5.14	2.63	5.66	2.85	4.82	2.42
5. Similarities	8.35	4.76	8.70	4.42	6.52	4.27	7.43	4.14
6. Vocabulary	19.89	8.11	19.45	7.08	17.99	7.44	19.65	7.13
7. Picture Arrangement	6.42	3.67	5.74	3.51	5.09	3.32	4.24	3.19
8. Picture Completion	9.06	3.02	7.17	2.67	7.68	3.28	6.75	2.60
9. Block Design	14.87	7.19	12.85	7.33	10.73	7.12	10.36	6.59
10. Object Assembly	15.09	4.78	13.67	5.12	12.84	5.25	12.79	5.18
11. Digit Symbol	25.15	11.81	24.85	12.30	18.14	9.45	19.40	10.55
12. WRAT Reading	71.02	25.20	77.03	22.03	68.57	24.90	73.12	26.35

diagnoses are positively related. The methods of measuring psychometric patterns are fairly adequate, though improvements will further crystallize the value of such measures and stimulate research that is likely to objectify the diagnosis. The psychiatric diagnosis as an external criterion of comparison has a large margin of error, but it can provide us with trends which are usable for normative purposes if cross validations are made to establish and explain the size and nature of the error. Two variables which interfere with ability patterning in diagnosis are age and sex. These variables should be controlled whenever rank analysis of subtests is undertaken. The most serious obstacles to the derivation of diagnostic ability ranks are distortions inherent in the norms of the Bellevue scale. Constant errors seem to push up some subtest scores and pull down others, and thereby obscure differences due to mental characteristics of the group. The number and nature of such errors can only be guessed at. Changes in difficulty levels since standardization, changes in materials, in directions of administration and scoring, biased population samplings in regard to personality traits, clerical errors, and varying proportions of sex and age samplings contributing to norms may be associated with the observed artifacts.

With respect to Wechsler's sign lists the following conclusions seem warranted:

1. The diagnostic signs hold up only when raw score means are used for purposes of comparison. Conversion of raw scores into weighted scores or quotients distorts subtest patterns.

2. His sign lists fit our female patterns more closely than they do our male patterns for each of the comparable groups.

3. Wechsler's neurotic, schizophrenic, and organic patients are inadequately differentiated from each other either in the psychiatric diagnosis or in the test results, or in both.

4. It is important that the patterns published in this paper be checked carefully on several large and independent groups of each diagnostic category to test both the reliability and validity of the ranks.

Received March 2, 1953.

References

1. Jastak, J. Problems of psychometric scatter analysis. *Psychol. Bull.*, 1949, **46**, 177-197.
2. Jastak, J. *Wide Range Achievement Test.* Wilmington, Del.: Story Co., 1946. Distributed by the Psychological Corporation, 522 Fifth Ave., New York 36, N. Y.
3. Rapaport, D., *et al. Diagnostic psychological testing.* Vol. I. Chicago: Year Book Publishers, 1945.
4. Wechsler, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.

Journal of Consulting Psychology
Vol. 17, No. 6, 1953

Sex Differences and Other Aspects of Young Superior Adult Performance on the Wechsler-Bellevue

Ralph D. Norman
The University of New Mexico

Interest in pathological performance on the Wechsler-Bellevue has been so great that, except for a few studies [2, 8], research with normal superior adults has been neglected. Sex differences have also received little attention, except for the item analysis by Jastak [6] and a recent note by Strange and Palmer [11]. It is the purpose of the present study to investigate with young superior adults a number of problems with clinical implications. These are sex differences as related to deterioration, verbal-performance differentials, clinical "signs," and scatter, as well as item difficulty and the problem of ceiling.

Subjects and Procedure

The subjects were 85 males and 68 females, volunteers for testing, with no known pathology. All cases in the author's files with a full scale IQ of 120 or better were used. According to Wechsler [12], they may be categorized as superior or very superior, constituting the topmost 8.9% of the standardization population. Age, so important in W-B scatter [5], was controlled so that no subject was younger than 15 or older than 29. Data on age were: males, range 18-29, mean 23.5; females, range 15-29, mean 22.3. The majority were college students, data on years of education being as follows: males, range 12-19, mean 14.4; females, range 10-18, mean 13.9. All tests were scored according to directions by Wechsler [12] or the supplementary manual by Kitzinger and Blumberg [7]. In order to test sex differences, it is important that full scale IQ's be reasonably equated. Table 1 presents data on IQ distributions, showing that matching was quite close. Full scale IQ statistics are: males, mean 127.5, *SD* 5.2; females, mean 126.8, *SD* 5.1. The *t* of the difference of 0.7 IQ points is statistically insignificant, being 0.91.

Table 1
Distributions of IQ's

IQ Group	Males		Females	
	N	%	*N*	%
120–124	28	32.9	27	39.7
125–129	28	32.9	20	29.4
130–134	18	21.2	15	22.1
135–139	9	10.6	4	5.9
140–144	2	2.4	2	2.9
Total	85	100.0	68	100.0

Results

Sex differences and deterioration. According to Table 2, there is a highly significant difference between superior adult males and females in Deterioration Quotient (DQ). The male mean DQ is 99.6, which is what is to be anticipated for their age group. The female DQ mean is 95.0, which is the DQ to be expected for ages 35-39 [12]. Thus our young, very bright females have a DQ which Wechsler says applies to an age group some 15 years older. As a matter of fact, 17.6% of these females have a DQ of 84 or less, which is what Wechsler gives as figure for ages 55-59, where-

Table 2
Quotients of Superior Adult Males and Females

Quotients	Males		Females			
	M	σ	M	σ	Diff.	*t*
Deterioration	99.6	10.7	95.0	10.7	4.6	2.61**
Verbal	126.3	7.0	122.7	6.7	3.6	3.19**
Performance	123.4	6.8	125.7	6.2	2.3	2.15*
Full scale IQ's	127.5	5.2	126.8	5.1	0.7	0.91

* Significant at the 5% level.
** Significant at the 1% level.

as only 5.9% of males have a DQ which is as low. The t of this difference of 11.7% is 2.18, significant at the 5% level.

Sex differences and verbal-performance differential. Not only is there an important sex difference in DQ, but Table 2 also reveals that there are highly significant sex differences in verbal IQ and performance IQ, although full scale IQ's are almost identical. Males are significantly higher in verbal (V), females in performance (P). There is also a significant difference between V and P *within* the sexes, the male VQ being 2.9 points higher than the male PQ, the t of this difference (using the r term) being 2.74, significant at the 1% level. The female VQ is 3.0 points higher than the female PQ, the t of this difference (using the r term) being 2.78, also significant at the 1% level. Wechsler states, "Subjects of superior intelligence generally do better on the verbal, and subjects of inferior intelligence do better on the performance part of the examination. There are also racial (group) and cultural differences" [12, p. 147]. It might be added that, according to evidence of this study, there are sex differences as well. The superior adult male does better in verbal, whereas the superior adult female does better in performance. Of interest here is that Strange and Palmer [11], reporting on unequated groups near average, also find males better in V than P, whereas females are better in P than V.

Wechsler [12] also states that for subjects with IQ's not far from average, a variation of 8 to 10 points between V and P in either direction is within normal range. Using his upper limit as important, we find that 24.8% of males, but only 5.8% of females have an excess of 10 or more points of V over P. This difference of 19.0% is highly significant, the t being 3.44, significant at the 1% level. Among males, 10.6% have an excess of 10 or more points of P over V, whereas 22.1% of females show this excess; the t of the difference of 11.5% is 1.88, significant between 5 and 10% levels.

Sex differences and subtests. Since DQ's, VQ's, and PQ's differ significantly between sexes, do outstanding subtest differences account for these results? Table 3 presents data for the 11 W-B subtests, demonstrating that males are significantly superior in Arithmetic, whereas females are significantly superior in Digit Symbol. The vastly greater inferiority of females in Arithmetic is what causes their lower DQ's, since Arithmetic is a "don't hold" test; it also contributes to their inferiority in VQ. Their superiority in Digit Symbol, plus slight superiorities in all other performance tests results in a significantly higher PQ. Jastak [5] says that males tend to be better in Information, Picture Completion, Object Assembly, Arithmetic, and Digit Span, whereas females tend to be better in Vocabulary, Digit Symbol, Comprehension, Picture Arrangement, and Block Design. According to Table 3, the present data are in agreement with Jas-

Table 3

Subtest Weighted Scores of Superior Adult Males and Females

Subtest	Males		Females			
	M	σ	M	σ	Diff.	t
Information	13.8	1.4	13.6	1.7	0.1	0.58
Comprehension	13.8	1.8	13.8	2.0	0.0	0.02
Digit Span	12.3	3.0	11.7	2.6	0.6	1.27
Arithmetic	14.2	2.5	11.8	2.8	2.4	5.54**
Similarities	14.9	1.8	14.9	1.5	0.0	0.12
Vocabulary	13.6	1.5	13.9	1.4	0.3	1.30
Pic. Arran.	12.9	2.5	13.4	2.6	0.5	1.22
Pic. Comp.	13.6	1.2	13.9	1.2	0.3	1.63
Block Design	14.3	1.8	14.7	1.7	0.4	1.36
Object Assem.	13.2	1.8	13.3	1.7	0.1	0.33
Digit Symbol	13.0	1.9	13.7	1.8	0.7	2.20*

* Significant at the 5% level.
* Significant at the 1% level.

Table 4

Significant Differences (*t*'s) between Means of Subtests of Superior Adult Males

	Information	Compr.	Digit Span	Arith.	Simil.	Vocab.	Pic. Arran.	Pic. Comp.	Block Design	Object Assem.
Comprehension	x	–								
Digit Span	4.16**	3.91**	–							
Arithmetic	x	x	*4.79***	–						
Similarities	*5.27***	*4.67***	*7.70***	*2.47**	–					
Vocabulary	x	x	*4.13***	x	6.50**	–				
Pic Arran.	2.71**	2.64**	x	3.19**	5.56**	2.13*	–			
Pic. Comp.	x	x	*3.81***	x	5.98**	x	*2.25**	–		
Block Design	*2.63***	*2.16**	*5.80***	x	2.77**	*4.89***	*4.35***	*4.07***	–	
Object Assem.	2.76**	x	*2.28**	2.75**	6.35**	x	x	x	6.15**	–
Digit Symbol	2.91**	4.31**	x	3.14**	6.68**	2.26*	x	2.44*	5.30**	x

x = no significant difference.
Numbers in italics indicate that row subtest means are greater than column means.
* Significant at the 5% level.
** Significant at the 1% level.

tak in all cases except Picture Completion, Object Assembly, and Comprehension.

Sex difference and "hard signs." Wechsler [12] states that for total scores beyond the limits 80-110, one should divide the mean subtest score by 4 to define a significant difference from the mean. This was done, and again the Arithmetic test strongly differentiated between the sexes. The data for significant deviations are as follows: positive—M, 8; F, 1; neither positive nor negative—M, 69; F, 45; negative—M, 8; F, 22. The χ^2 of these differences is 15.34, significant at the 1% level. Females apparently have a larger number of negative, males of positive significant deviations.

Sex differences and intertest scatter. Examination of Table 3 indicates that there is considerable variability *within* sexes in terms of subtest means. Since the assumption in clinical usage of the W-B is that in normal adults there should be comparatively little variation among subtest scores, it is important to note whether high intelligence and sex make appreciable differences in subtest means. Accordingly, significance tests were run (using the *t* terms), between subtest means for each sex separately. Table 4 yields results for superior adult males, and Table 5 for females. These tables are noteworthy for the very high number (far beyond chance) of statistically significant differences which are obtained between certain subtests and others.

Typical superior adult male scatter would not only indicate V greater than P, as noted above, but

Table 5

Significant Differences (*t*'s) between Means of Subtests of Superior Adult Females

	Information	Compr.	Digit Span	Arith.	Simil.	Vocab.	Pic. Arran.	Pic. Comp.	Block Design	Object Assem.
Comprehension	x	–								
Digit Span	5.33**	5.16**	–							
Arithmetic	4.68**	4.83**	x	–						
Similarities	*5.75***	*3.56***	*8.17***	*8.50***	–					
Vocabulary	x	x	*6.32***	*6.25***	5.14**	–				
Pic. Arran.	x	x	*3.73***	*3.51***	3.97**	x	–			
Pic. Comp.	x	x	*6.67***	*5.46***	4.42**	x	x	–		
Block Design	*3.96***	*2.86***	*8.32***	*7.96***	x	*2.99***	*3.14***	*3.06***	–	
Object Assem.	x	x	*3.99***	*3.99***	6.49**	2.59**	x	2.68**	4.96**	–
Digit Symbol	x	x	*6.04***	*5.00***	3.91**	x	x	x	3.50**	x

x = no significant difference.
Numbers in italics indicate that row subtest means are greater than column means.
* Significant at the 5% level.
** Significant at the 1% level.

also the following (Table 4): Digit Span significantly lower than 8 other subtests (all except Picture Arrangement and Digit Symbol); Similarities significantly higher than all 10 other subtests; Picture Arrangement significantly lower than 7 others (all except Digit Span, Object Assembly, and Digit Symbol); Block Design significantly higher than 8 others (all except Arithmetic and Similarities); and Digit Symbol significantly lower than 7 others (all except Digit Span, Picture Arrangement, and Object Assembly). These results are comparable with studies by Estes [2] and Merrill and Heathers [8] on other college males. Using deviations from Vocabulary, Estes found negative deviations in Digit Span, Picture Arrangement, and Digit Symbol, positive deviation in Similarities, and zero deviation in Block Design. With the same scatter measure, Merril and Heathers found negative deviations in Digit Span and Digit Symbol, positive deviations in Similarities and Block Design, and practically zero deviation in Picture Arrangement.

Typical superior adult female scatter is even more pronounced. P is not only greater than V, but the following are also shown (Table 5): Digit Span significantly lower than all other subtests except Arithmetic; Arithmetic significantly lower than all others except Digit Span; Similarities significantly higher than all others except Block De-

Table 6

Significant Sex Differences in Items for Superior Adults

Test	Item	% Success M	% Success F	Diff. (%)	*t*	Superiority	Agrees with Jastak [6]
Information	15. Population	81.1	53.0	28.1	3.77**	M	Yes
	16. Wash. birth.	57.6	75.0	17.4	2.28*	F	Yes
	18. Egypt	95.3	85.3	10.0	2.00*	M	Not found
Arithmetic	Problem 6	98.8	86.8	12.0	2.82**	M	Not found
	Problem 8	89.4	75.0	14.4	2.29*	M	Yes
	Problem 9					M	Yes
	(0 sols.)	29.4	51.5	22.1	2.80**		
	(2 + sols.)	55.3	28.0	27.3	3.54**		
	(3 sols.)	20.0	8.8	11.2	1.99*		
	Problem 10					M	Yes
	(3 sols.)	29.4	58.8	29.4	3.77**		
	(2 + sols.)	50.6	28.0	22.6	3.02**		
	(3 sols.)	23.5	5.9	17.6	3.20**		
Similarities	10. Poem-Statue (2 sols.)	61.2	78.0	16.8	2.28*	F	Yes
Vocabulary	23. Vesper	73.1	87.3	14.2	2.12*	F	Yes
	28. Ballast	84.6	60.3	24.3	3.28**	M	Yes
	30. Spangle	59.0	87.3	28.3	3.99**	F	Yes
Pic. Arran.	4. Flirt					F	Yes
	(3 sols.)	71.7	88.3	16.6	2.61**		
	(JANET)	36.5	54.4	17.9	2.25*		
	(AJNET)	22.3	10.3	12.0	2.08*		
Pic. Comp.	11. Image	81.1	92.7	11.6	2.18*	F	Not found
	13. Thread	81.1	63.3	17.8	2.45*	M	Yes
	14. Brow	50.6	85.3	34.7	4.96**	F	Yes
Block Design	Item 2 (6 sols.)	7.1	19.1	12.0	2.17*	F	Yes
	Item 4 (4 + sols.)	74.1	91.2	17.1	2.88**	F	Yes
Object Assem.	Profile (10 sols.)	14.1	38.3	24.2	3.45**	F	Not found

* Significant at the 5% level.
** Significant at the 1% level.

sign; and Block Design significantly higher than all others, except Similarities. Taking both sexes together, Similarities and Block Design are significantly high while Digit Span is significantly low.

Sex differences in item success. Jastak [6] suggests that item analyses for sex differences be made by others either to confirm or to deny his results. Table 6 gives data on significant superior adult sex differences in various scale items. Also indicated is whether results agree generally with Jastak.[1]

Again, there is strong male superiority in Arithmetic items, especially problems 9 and 10. For the scale as a whole, 10 items favor females, 8 favor males. This difference is not as overwhelming as that found by Jastak (32 favoring females, 18 males). Generally though, where we have found significant differences, we agree with him. We find results he did not, in that more males than females succeed with "Egypt" in Information and with problem 6 in Arithmetic, while more females succeed with "Image" in Picture Completion, and "Profile" in Object Assembly. An interesting difference not mentioned by Jastak is that a significantly greater percentage of females give JANET solutions to "Flirt" in Picture Arrangement, whereas a significantly greater percentage of males give AJNET solutions. Males are penalized here since the former solution carries 3 points of credit, the latter 2 points.

A word might be said about the Digit Span test. Jastak [6] indicates that this test is loaded with female superiority. He finds, e.g., that females are significantly better in Forward 7, and in Backward 4, 5, 6, and 7. We find no such differences, and if anything, slight but consistent superiority for males.

Jastak states, too, that females are quicker in executing Block Designs. We find this to be true. The mean number of points earned by females is greater in 5 of the 7 problems (all except problems 3 and 6) and they are superior in the types of solutions for problems 2 and 4, as given in Table 6.

Order of difficulty of items. Order of difficulty is, of course, not directly comparable between our superior adults and either Wechsler's [12] or Jastak's [6] group. However, interest attaches to which items cause particular difficulty in a highly intelligent group, since less intelligent groups would generally also experience such difficulty. Over-all success was calculated by averaging male and female successes with each item.

In the Information test, the following items are relatively harder for our superior subjects than their regular positions indicate: "Pints" (our rank order 13, regular rank order 5); "Height" (ours 16, regular 9); "Paris" (ours 17, regular 12); and "Pole" (ours 22, regular 17). The following are relatively easier: "Brazil" (ours 5.5, regular 11); "Hamlet" (ours 5.5, regular 14); "H. Finn" (ours 12, regular 19); and "Vatican" (ours 15, regular 20). The last 5 items in the test are apparently approximately correctly placed. Jastak makes the "Pints" question number 4 in order, but we suspect it is more difficult than this. He leaves "Height" and "Hamlet" unchanged, moves "Paris" and "Pole" down and "Brazil" and "H. Finn" and "Vatican" up. The strong influence of schooling is shown in our group in that "H. Finn" and "Hamlet" are easier than "Pints" or "Height." Our ranking correlates .855 with Jastak, .834 with Wechsler.

In Comprehension, scoring 0, 1, or 2 credits, and taking mean points, we find "Shoes" misplaced for our superior adults, seemingly being difficult (our rank 10, regular 5). Intelligent subjects find it hard to give 3 reasons in order to earn 2 points on this item. If Wechsler persists in this type of scoring he should move the item farther down. Eglash [1] also mentions this problem with "Shoes." The "Deaf" item is relatively easy for our group (our rank 5, regular 10), which may be accounted for by superior education, the "learning" criterion for 2 points being quickly achieved. We do not entirely agree with Jastak when he says that Comprehension has a dearth of items at upper levels of difficulty. His method of scoring success (i.e., any success) indicates this; our adherence to point values does not since "Shoes," "Laws," and "Marriage" are closer to a mean of 1 than 2 points for our subjects. The ρ's are .479 with Jastak, .645 with Wechsler.

Using mean point values again, we find that in Similarities the "Air-Water" and "Wood-Alcohol" items are relatively harder for superior adults; "Poem-Statue" is relatively easier. The former rank order 10 and 12 respectively as against regular positions of 6 and 7; the latter ranks 6, regularly 10. Jastak moves "Wood-Alcohol" down, but leaves the others unchanged. Our rank order correlates .727 with Jastak, .671 with Wechsler.

In Arithmetic, practically 100% success is achieved for the first 5 problems and for problem 7 by our group. Problem 6 is more difficult for females, and if both sexes are considered, ranks after problem 7. We thus agree with Jastak on the greater difficulty of problem 7 over problem 6. Problems 9 and 10 are of about equivalent difficulty for males, but not for females, the 10th being harder. The ρ's are .936 with Jastak, .864 with Wechsler.

---

[1] Our data are not directly comparable with Jastak's for several reasons: (*a*) we have a much smaller N (153 vs. 1,176); (*b*) we used only 120 IQ or better, whereas Jastak used a much wider IQ range; (*c*) we made a number of statistical tests not performed by Jastak, who defined success with an item as *any* success, whereas we tested for differences at different *levels* of success; (*d*) Jastak gives no indication of having matched his IQ's as carefully as we did. For example, 26% of his females were nurses, who might be presumed to be of above average intelligence.

On the Vocabulary test, the following are relatively harder for our group: "Fable" (our rank order 20.5, regular 15); "Brim" (ours 20.5, regular 16); "Armory" (ours 23, regular 14); "Vesper" (ours 29, regular 23); "Ballast" (ours 32, regular 28); "Dilatory" (ours 40, regular 36); and "Amanuensis" (ours 41, regular 37). The following are relatively easier: "Plural" (ours 8, regular 18); "Nitroglycerine" (ours 8, regular 20); "Microscope" (ours 8, regular 22); "Hara-kiri" (ours 22, regular 34); "Aseptic" (ours 36, regular 40). We are in almost complete agreement with Jastak who has moved all our "harder" words (except "Brim" and "Dilatory") down, and all our "easier" words (except "Plural") up. Certainly words like "Dilatory" and "Amanuensis," on which only 16.4% and 6.0% of our group are successful, belong down at the very bottom of the list right before "Traduce." The ρ's are .933 with Jastak, .917 with Wechsler.

For Picture Arrangement, order of difficulty is as Wechsler has them. Practically 100% success is achieved on the first 3 items by our superior adults. Items 4, 5, and 6 are sufficiently difficult for them, since 80.0, 23.5, and 10.6%, respectively, achieve perfect solutions.

The Picture Completion test again reveals more agreement with Jastak than Wechsler. "Diamond" and "Leg" each rank 10.5 in our group, but 4 and 5 in Wechsler's regular order; "Stacks" is also relatively harder for superior adults, ranking 13 as against regular rank 7. Items which are easier for our group are "Hand," (ours 4, regular 9); "Water" (ours 2, regular 10); and "Shadow" (ours 8, regular 15). Jastak moves all our harder items down, and all the easier (except "Water" which he leaves unchanged) up. The correlations are .785 with Jastak, .517 with Wechsler.

In Block Design, the 4th item is especially difficult for our superior adults, but the 5th and 7th are easier. Rank order of the 7 items in our group is 1, 3, 4, 6, 2, 7, 5. Jastak agrees substantially, moving No. 4 down, No. 5 up, but leaving No. 7 unchanged. It would seem that some revamping of time credits is needed in this test if the present order is maintained. The ρ's are .786 with Jastak, .643 with Wechsler.

Over all, we find that Jastak [6] is essentially in agreement in his item rearrangement, even though we base conclusions on a highly selected group. Our ρ's with Jastak are higher than our ρ's with Wechsler in Information, Similarities, Arithmetic, Vocabulary, Picture Completion, and Block Design, and higher with Wechsler only in Comprehension.

Ceiling of subtests. A unique opportunity is afforded when studying superior IQ's to examine discriminative ceilings of subtests. With 153 subjects, we contrasted lowest and highest quarters, or IQ's of 120–123 and 131–144 respectively. The lower group N was 44 (to include all 11 subjects of 123 IQ), with 24 males (54.5%) and 20 females (45.5%). The upper group N was 38, with 23 males (60.5%) and 15 females (39.5%); they are definitely in Wechsler's [12] very superior classification. The two groups thus have comparable sex percentages. Table 7 presents com-

Table 7

Comparative Subtest Ceilings

Lower vs. Higher Superior Adult IQ's

Subtest	IQ 120–123 ($N = 44$)		IQ 131–144 ($N = 38$)			
	M_{wts}	σ	M_{wts}	σ	Diff.	t
Information	13.0	1.4	14.5	3.2	1.5	2.75**
Comprehension	13.0	1.6	14.6	1.8	1.5	3.97**
Digit Span	10.6	2.2	13.8	2.2	3.2	6.44**
Arithmetic	12.1	2.1	14.8	2.5	2.7	5.24**
Similarities	14.3	1.6	16.1	1.3	1.8	5.53**
Vocabulary	13.1	1.4	14.8	1.2	1.7	5.71**
Pic. Arran.	12.5	2.4	13.9	2.7	1.4	2.46*
Pic. Comp.	13.4	1.1	14.1	0.9	0.7	3.13**
Block Design	13.5	1.6	15.6	1.4	2.2	6.46**
Object Assem.	12.5	1.8	14.3	1.2	1.8	5.18**
Digit Symbol	12.6	1.8	13.9	1.8	1.3	3.19**

* Significant at the 5% level.

** Significant at the 1% level.

parative subtest ceilings, contrasting both groups.

According to Table 7, Block Design, with a t of 6.46, is the best discriminating test. Wechsler holds it in high respect, saying that "it conforms to all statistical criteria for a good test" [12, p. 97], and reporting that it correlates very well with the scale as a whole. Digit Span (t of 6.44) ranks next in discriminatory power, surprisingly, since Wechsler says, "Memory for digits correlates with general intelligence up to a certain point, but beyond the ability to repeat 6 or 7 digits forward, its correlation becomes negligible" [12, p. 74] and also, "A good rote memory is of practical value but correlates very little with higher levels of intelligence" [12, p. 85]. We cannot agree with him since the test still is sharply discriminating between superior and very superior intelligence. Vocabulary, Similarities, and Arithmetic (t's of 5.71, 5.53, and 5.24 respectively) hold up as discriminators quite well, and we agree with Wechsler who holds them in high esteem. Object Assembly, like Digit Span, is somewhat surprising (t of 5.18) since Wechsler feels that formboard-type tests "at the high levels have very little discriminative value" [12, p. 97]. Two other tests Wechsler rates lowly as measures of general intelligence are Picture Completion and Picture Arrangement. He says that Picture Completion is relatively inadequate in discriminating between higher levels of intelligence (i.e., high average and superior). We find a comparatively low t of 3.13 for this test, although it should be mentioned that a low ceiling was built into it by Wechsler himself since no more than 15 points may be earned on it. Our lowest t is for Picture Arrangement (2.46), which Wechsler reports as correlating poorly with other tests and with the scale as a whole. Information, a test Wechsler respects, is also comparatively low (t of 2.75), but it may not be as poor as it seems in view of the high degree of education in both our contrasted groups.

Generally, however, although t's between a superior and very superior group vary widely, all tests retain their discriminatory powers quite well, even those about which Wechsler has misgivings.

Discussion

Clinical implications of this study are several. Sex differences are great enough to cause significant discrepancies in DQ's and verbal-performance differentials, as well as in "hard signs." The clinician examining the older, brighter female may not know how much deterioration is a function of age or how much of sex, since in our young, maximally functioning females a loss equivalent to 15 years of age is already apparent in DQ. The V-P differential, of which Wechsler makes much as a clinically important sign, is also affected by sex differences, as is the "hard sign" for Arithmetic. Also, typical scatter patterns exist for both superior females and males, independent of pathology. Other clinicians, such as Garfield [4], Schofield [9], French and Hunt [3], and Schnadt [10], have already mentioned the import of factors outside of pathology as instrumental in scatter. Even with our highly selected group, there is strong suspicion that item difficulty is not quite as Wechsler has given it, since all but one of our rank orders correlate higher with Jastak than with Wechsler. Jastak [6] feels that accuracy in item scaling is essential in order to derive clinically meaningful measures of intratest scattering, as well as to provide for economy, ease, reasonable limits, and accuracy of testing. The matter of ceiling is important for determination of discriminative values of tests between higher intellectual levels. As noted above, tests like Digit Span and Object Assembly are perhaps not as poor high level discriminators as Wechsler believes.

Summary

Using a group of young adults, all with 120 IQ or better on the Wechsler-Bellevue, and with sexes matched in full scale IQ, it was found that:

1. Significant differences exist between sexes in DQ, VQ, and PQ, and within sexes in VQ and PQ. Young, superior females appear to be more "deteriorated" than their age warrants. They are significantly inferior to males in Arithmetic, in which they also have significantly more low deviant "hard signs."

2. Other sex differences in items appear which affect the clinical utility of the test, and which generally corroborate previous research.

3. Highly significant scatter patterns exist which are functions of both sex and high IQ level.

4. Study of order of difficulty of items of subtests generally supports Jastak's [6] contention that there are important discrepancies in item gradation.

5. All subtests discriminate between Su-

perior (IQ 120—123) and Very Superior (IQ 131—144) groups, but not to the same degree.

Received February 13, 1953.

References

1. Eglash, A. Validation of the Wechsler "shoes" item. *J. abnorm. soc. Psychol.*, 1950, **45**, 733-734.
2. Estes, S. G. Deviations of Wechsler-Bellevue subtest scores from vocabulary level in superior adults. *J. abnorm. soc. Psychol.*, 1946, **41**, 226–228.
3. French, Elizabeth G., & Hunt, W. A. The relationship of scatter in test performance to intelligence level. *J. clin. Psychol.*, 1951, **7**, 95–98.
4. Garfield, S. L. An evaluation of Wechsler-Bellevue patterns in schizophrenia. *J. consult. Psychol.*, 1949, **13**, 279–287.
5. Jastak, J. Problems of psychometric scatter analysis. *Psychol. Bull.*, 1949, **46**, 177-197.
6. Jastak, J. An item analysis of the Wechsler-Bellevue tests. *J. consult. Psychol.*, 1950, **14**, 88–94.
7. Kitzinger, Helen, & Blumberg, E. Supplementary guide for administering and scoring the Wechsler-Bellevue Intelligence Scale (Form I). *Psychol. Monogr.*, 1951, **65**, No. 2 (Whole No. 319).
8. Merrill, R. M., & Heathers, Louise B. Deviations of Wechsler-Bellevue subtest scores from vocabulary level in university counseling-center clients. *J. consult. Psychol.*, 1952, **16**, 469–472.
9. Schofield, W. Critique of scatter and profile analysis of psychometric data. *J. clin. Psychol.* 1952, **8**, 16-22.
10. Schnadt, F. Certain aspects of Wechsler-Bellevue scatter at low IQ levels. *J. consult. Psychol.*, 1952, **16**, 456–461.
11. Strange, F. B., & Palmer, J. O. A note on sex differences on the Wechsler-Bellevue tests. *J. clin. Psychol.*, 1953, **9**, 85–87.
12. Wechsler, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.

Journal of Consulting Psychology
Vol. 17, No. 6, 1953

Why Wechsler-Bellevue Full-Scale IQ's Are More Variable Than Averages of Verbal and Performance IQ's

Julian C. Stanley [1]
University of Wisconsin

Many clinicians have been puzzled by discrepancies similar to those in Table 1.[2] For persons whose Verbal and Performance IQ's on Form I of the Wechsler-Bellevue Intelligence Scale average 130, the Full-Scale IQ is greater than 130, and for those whose *V* and *P* IQ's average 70 the *FS* IQ is less than 70. That this result for IQ's differing much from 100 is a mathematical necessity rather than an indictment of Wechsler's tables has been sensed by most psychologists, but so far as the writer knows no explanation has been published.

Table 1
Wechsler-Bellevue I Data [2] Showing That Full-Scale IQ's Are Not Merely Averages of Verbal and Performance IQ's

Age in Years	Verbal IQ	Performance IQ	Average of Verbal and Performance IQ's	Full-Scale IQ
10.0	120	140	130	133
20-24	130	130	130	133
45-49	70	70	70	67

The phenomenon occurs with all kinds of averaged measurements that are not perfectly intercorrelated. The teacher who gives several tests during a quarter and then averages the marks each pupil has received notices that these final marks include fewer *A*'s and *F*'s than he customarily assigns to the separate tests. The more tests he administers, the greater will be this "shrinkage." Also, the less reliable and/or related his quizzes are the fewer high or low final averages will his students obtain. Obviously, distributions of means are virtually always less variable than the single-score distributions from which they were obtained. Usually the teacher merely redistributes the average marks so that his customary percentages of letter grades are given. This procedure is quite proper in most instances, as the teacher is probably interested in only the relative standings of the students within his own group. There are two cases, however, when the teacher should be aware of the effect of averaging:

1. Average scores may be interpreted as belonging to a distribution of scores with a "standard" standard deviation. For example, IQ scores are used to identify a person's performance with respect to the performance of others. We say that an individual who attains an IQ of 130 has obtained a score higher than that secured by approximately 97 or 98 per cent of the people in the standardizing group. We so interpret the score because IQ's are usually scaled to yield a standard deviation of 15-16½. If several IQ's for the same individual are averaged, the mean score is *not* an IQ, since such means have a smaller standard deviation, depending upon the number of scores averaged and the intercorrelations of the tests used. The error will not usually be serious if the intercorrelations are high.

2. Average scores may be misinterpreted when not all the individuals to be ranked have the same number of scores. If there is only one mark for a certain student and several for all of the other students, that student's mark should be regressed toward the rest of the class's general mean before it is included with the averaged marks in the distribution used to find the final ranking of the students. This is likely to be a practically impossible task for the usual teacher, however. Here again, the

---

[1] The writer is greatly indebted to Dr. Jerome P. Doppelt and Mr. Robert F. Schweiker for many helpful suggestions.

[2] See, for example, the comments of Sloan and Schneider [1].

seriousness of the error depends on the number of scores averaged and the intercorrelations of the tests.

The implications of this restricted variability become obvious to anyone who plots on a profile several high or low local-norm standard scores and the standard score based upon their raw-score sum. If the testee did poorly on all tests, his summary standard score will be substantially below the average of his separate standard scores—perhaps lower than any one of them. If he did well, his summary score will be higher than the average of the separate standard scores. The more tests involved, the greater these discrepancies are likely to be.

Technical Note

It can be shown that when each of *N* individuals has *n* different test scores, the variance of the distribution of the *N* average scores is equal to 1/*n*th of the sum of the average variance of the *n* tests and $(n-1)$ times the mean of the $\frac{1}{2}n(n-1)$ different covariances among the *n* tests:

$$\sigma_M{}^2 = \frac{1}{n}\left[\overline{\sigma^2} + (n-1)\overline{r_{ij}\sigma_i\sigma_j}\right] \qquad i \neq j$$

$$= \frac{\overline{\sigma^2}}{n} + \frac{n-1}{n}\overline{r_{ij}\sigma_i\sigma_j} \qquad (1)$$

This formula is the general basis for the various statements already made. If there are only two tests $(n=2)$, $\sigma_M{}^2$ becomes $\frac{1}{4}(\sigma_1{}^2 + \sigma_2{}^2 + 2r_{12}\sigma_1\sigma_2)$. When dealing with the Wechsler-Bellevue Intelligence Scales, Forms I or II, for which the standard deviation of Verbal or Performance IQ's at any age level is 14.83 and the correlation between *V* and *P* IQ's is about .62, the formula yields a standard deviation of 0.9(14.83), or 13.35, a "shrinkage" of 10 per cent. Taking the Full-Scale IQ of 133 from Table 1, we note that 100 + 0.9(133 − 100) equals 130, the average of *V* and *P* IQ's shown there.

It can be proved that the correlation between Full-Scale IQ's and the average of Verbal and Performance IQ's is

$$\frac{(\sigma_v + \sigma_p)\sqrt{\dfrac{1 + r_{vp}}{2}}}{\sqrt{\sigma_v{}^2 + \sigma_p{}^2 + 2r_{vp}\sigma_v\sigma_p}},$$

where *v* and *p* subscripts refer to Verbal and Performance weighted scores (*not* IQ's) at the particular age level under consideration. Probably σ_p will be somewhat smaller than σ_v, since the intercorrelations of Performance tests are in general less than those of Verbal tests. If σ_v and σ_p were identical or if r_{vp} were 1, the formula would equal 1, its maximum possible value.

As *n*, the number of tests, approaches infinity, $\sigma_M{}^2$ in Formula (1) approaches $\overline{r_{ij}\sigma_i\sigma_j}$, the average covariance. If, however, $\overline{r_{ij}\sigma_i\sigma_j}$ is zero, $\sigma_M{}^2$ equals $\overline{\sigma^2}/n$, and as *n* becomes large, the variance of the means rapidly gets smaller.

Summary

Some clinicians have wondered why the average of Wechsler-Bellevue Verbal and Performance IQ's differs in most instances from the Full-Scale IQ, the former being 10 per cent nearer 100. Average and *FS* IQ's do not correlate perfectly, so nothing is gained by using such averages, which are not IQ's in the usual sense.

Reduced variability of individuals' means is common to all kinds of measurements.

Received February 10, 1953.

References

1. Sloan, W., & Schneider, B. A study of the Wechsler Intelligence Scale for Children with mental defectives. *Amer. J. ment. Def.*, 1951, 55, 573–575.
2. Wechsler, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.

Journal of Consulting Psychology
Vol. 17, No. 6, 1953

The Significance of Artistic Excellence in the Judgment of Adjustment Inferred from Human Figure Drawings[1, 2]

John W. Whitmyre
Veterans Administration Hospital, Salt Lake City, Utah

Psychologists who utilize the human figure drawing as a projective instrument are often asked whether their inferences about the level of personal adjustment judged from the drawings are not basically inferences concerning the degree of artistic excellence which the drawings possess. A review of the literature has revealed no studies aimed directly at determining the answer to this question, and to find information concerning this problem is the primary aim of this study.

While many human figure drawing investigations have dealt with the drawings detail by detail, Albee and Hamlin [1] have recognized that this is not the technique utilized by clinicians in their daily use or misuse of the human figure drawing. They state that the aim of their investigation was to find whether "global or insightful impressions by clinicians who are experienced with projective methods have significant reliability and validity" [1, p. 390]. Their method was to permit experienced psychologists to rank a series of human figure drawings according to the level of personal adjustment which the judges felt was reflected in each drawing as a totality. They found exceedingly reliable interjudge judgments could be made, as well as rather valid ones, with the criterion of adjustment being arrived at from case history material. The investigation was expressly designed to avoid evaluating the drawings piecemeal. It is the belief that projective test protocols are best approached from the configurational point of view which led to the plan of the present study.

Statement of Hypotheses

In order to evaluate the extent to which the artistic merit of human figure drawings is related to the inferences made about the personal adjustment of the artist, the following hypothesis was tested.

Hypothesis I. There is no significant relationship between the degree of artistic excellence reflected in drawings of human figures and level of personal adjustment as judged by psychologists from human figure drawings.

In order to learn whether psychologists are able to judge human figure drawings differently when evaluating them in terms of artistic merit rather than in terms of personal adjustment, a second hypothesis was tested.

Hypothesis II. Clinical psychologists judge human figure drawings differently when evaluating them in terms of level of personal adjustment reflected therein, than when evaluating them in terms of degree of artistic excellence.

Since the value of any psychological test hinges ultimately upon its validity, the following hypothesis was also tested.

Hypothesis III. Actual level of personal adjustment is more closely related to level of personal adjustment as judged from drawings of human figures than to level of artistic excellence which the drawings are judged to possess.

Procedure

Fifty drawings of a man and a woman, both on the same sheet of paper, were collected from Veterans Administration psychological files. All psychiatric patients who had executed these drawings were between 22 and 33 years of age at the time of testing and were white, male, veterans of World War II who had achieved Wechsler-Bellevue IQ's of at least 100. All had completed at least eight grades of schooling. No patients with known central nervous system damage were used. Fifty human figure drawings were then collected from veterans functioning adequately in daily life and without diagnosis or history of neuropsychiatric difficulties, but meeting the same requirements as the patient sample on the variables of age, race, veteran status, and educational achievement. In neither group were fragmentary drawings utilized.

---

[1] Submitted to the Graduate School of the University of Pittsburgh in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

[2] From the Veterans Administration Regional Office, Pittsburgh, Pa.

Two sets of drawings were then compiled, each set consisting of 25 psychiatric and 25 nonpsychiatric cases. In each drawing group, the psychiatric and nonpsychiatric subjects whose drawings made up that drawing group were equated in terms of mean age and education.

After the obliteration of all identifying information, the above-described groups of drawings were presented to a group of eight advanced commercial art students for ranking according to the artistic merits of the drawings. The drawings were also presented to two six-man groups of clinical psychologists holding Ph. D. degrees and having two to six years' experience with the human figure drawing as a projective test. These two groups of psychologists were equated for experience with the human figure drawing. Where one group of psychologists ranked a set of drawings for artistic merit, the other group ranked the same drawings for adjustment reflected, and vice versa. All drawings were thus evaluated for both art and adjustment by the psychologists. Each psychologist used his own concept of what constitutes "adjustment" as it is commonly used by clinicians. In all cases, counterbalancing for order of presentation was observed. It was thus possible to procure artists' art evaluations, psychologists' art evaluations, and psychologists' adjustment evaluations of all of the drawings in each group.

In all cases, the rankings which the judges assigned to the drawings were converted to normal-curve base-line deviates, or scale values, according to the procedure outlined by Guilford [3]. In order to find the values appropriate to the testing of the hypotheses of the investigation, correlational procedures were carried out utilizing the scale values assigned the drawings by the artists and psychologists. To find the reliability values, each of the judging groups was divided into two equal subgroups, the scale values assigned each drawing by the subgroups determined, and the subgroup scale values correlated.

Results

Reliability. In all judging groups exceedingly reliable results were found. Table 1 summarizes the reliability figures. In each case the reliability coefficient was found by dividing the judging group in half and correlating the scale values assigned the drawings by these subgroups. All these correlation coefficients were then expanded by the Spearman-Brown prophecy formula to correct for halving the number of judges to compute the correlation.

Table 1
Reliability Coefficients and Standard Errors for All Judgments Corrected by the Spearman-Brown Formula

Judges	Drawing group	Judged for	r	SE
Artists	I	Art	.943	.015
Artists	II	Art	.936	.018
Psychologists II	I	Art	.936	.018
Psychologists I	II	Art	.969	.009
Psychologists I	I	Adjustment	.884	.033
Psychologists II	II	Adjustment	.937	.018

By the use of Fisher's r-into-z transformation described by Edwards [2], it was possible to determine whether any of these reliability values differ significantly. The judgments yielded by the psychologists when evaluating the drawings in terms of adjustment are slightly, yet significantly, less reliable than the other two judgments, psychologists' judgments for art and artists' judgments for art. The difference is significant at the .05 level of confidence.

Hypothesis I. This hypothesis was tested by the correlations between the psychologists' adjustment scale values assigned the drawings and the artists' art scale values assigned the drawings. For one set of drawings, the value was found to be .777 ± .058 and for the other set, .792 ± .053. These significant, consistent, and moderately high correlations indicate rejection of the hypothesis which states that no significant relationship exists between degree of artistic excellence reflected in drawings of human figures and level of personal adjustment inferred by psychologists from these drawings. It is apparent that the degree of artistic excellence which a drawing possesses is highly related to the degree of personal adjustment which the drawing is judged to reflect.

Hypothesis II. This hypothesis was evaluated with the correlations between adjustment scale values assigned a set of drawings by one group of psychologist judges and the art scale values assigned the same set of drawings by the other, equated group of psychologist judges. With one set of drawings, the value found was .885 ± .039. The correlation for the other set of drawings was .859 ± .037. These high, significant, and similar coefficients of correlation indicate rejection of the hypothesis which holds that psychologists judge human figure drawings differently when evaluating them in terms of level of personal adjustment than when evaluating them in terms of artistic excellence. Rather, they seem to have judged the drawings in much the same manner in each case. They apparently attended essentially to the artistic merits of the drawings. This impression is supported by application of partial correlation procedures to determine the correlation between the art judges' evaluations of the drawings for art and the psychologists' evaluations of the drawings for adjustment with the effects of the psychologists' art evaluations "held constant," or partialed out. In the case of each set of drawings, the high correlations between the psychologists' evaluations of adjustment and the artists' evaluations of art fell to insignificance when the psychologists' art evaluations were partialed out.

Hypothesis III. To test this hypothesis, the pertinent values were the point-biserial correlations between the artists' art judgments and the criterion of adjustment, and between the psychologists' adjustment judgments and the same criterion. In one case, the correlation between the artists' scale values and the dichotomous criterion psychiatric patient *vs.* nonpsychiatric subject was .095. The psychologists' adjustment values in this case correlated .237 with the criterion. Neither of these values is significantly different from zero. In the case of the other set of drawings, the artists' values correlated .101 with the criterion, while the psychologists' adjustment values correlated .179 with the criterion. Again, these values are not significant.

It was possible, utilizing the partial correlation method, to find values representative of the correlation between the psychologists' adjustment values and the criterion dichotomy with the influence of the art values assigned to the group by the psychologists partialed out. In the first instance, this procedure resulted in an increase in the correlation from .273 to .372, which is a value significant at the .01 level of confidence. In the second instance, the value of .169 was reduced to .002 when the influence of the psychologists' art ratings was partialed out. Hypothesis III must, then, be rejected, for the correlations indicate that actual level of personal adjustment is no more closely related to level of adjustment as judged from drawings of human figures than to level of artistic excellence which the drawings are judged to possess. Also, the influence of the psychologists' unconsciously judging according to artistic merit in one instance raised, and in the other instance lowered, the correlation between their adjustment ratings of the drawings and the criterion of adjustment.

Discussion

Perhaps the most clear-cut finding of this investigation is that there is a significant relationship between level of personal adjustment judged to be reflected in human figure drawings and the degree of artistic merit which they are judged to possess. Without regard to the validity or lack of validity of the human figure drawing as a clinical instrument, it appears that the psychologists who served in this investigation ranked the drawings in such a way that their ratings corresponded to a marked extent with the art evaluations by the artists. Furthermore, the high correlations between the art and adjustment evaluations made by the two groups of equated psychologist judges indicate that, to a marked extent, they evaluate the drawings in much the same manner, whether they set out to rate art or adjustment.

It must be remembered that in each case the psychologists who rendered the art values for a given group of drawings were not the same psychologists who rendered the adjustment values for those same drawings. Here the assumption that these two groups of psychologist judges are equal in ability to deal with the figure drawings is a most basic and necessary one. The similarity of the reliability and other comparable correlational values suggests that this assumption was a safe one.

Furthermore, it appears that neither art nor adjustment ratings show any appreciable relationship to the criterion of adjustment, patient or nonpatient status. If one accepts this rough criterion of personal adjustment, one must account for the failure of the ratings of these "average" clinical psychologists to distinguish the two differently adjusted groups. Observation reveals that many of the drawings by "normal" veterans are fully as inadequate as many seen in NP hospital files. This is consistent with many of the findings of Roe [4, p. 336], who has done extensive work involving projective tests with normals. She has found that many disturbed projective test protocols are yielded by apparently well-functioning individuals. Her interpretation is that while the basic difficulties as projected on tests may be the same in the adjusted and maladjusted, the former group has faced and dealt with the difficulties while the latter group has not [4, p. 339].

If it is true that underlying personality structure can be similar in two persons, one well-adjusted by society's standards and one inadequately adjusted, then it appears that this underlying structure may be the feature of individual make-up which projective tests tap. If it is not possible for the average clinical psychologist to differentiate those who are well-adjusted from those who are not on the basis of the human figure drawing, it appears that use of the instrument to assay level of adjustment is questionable. Either more adequate and communicable means of determining level of adjustment on the basis of drawings is necessary, or else psychologists should perhaps limit themselves to drawing inferences about features other than adjustment level from the drawings.

Summary and Conclusions

Human figure drawings were collected from psychiatric patients and "normal" veterans. Clinical psychologists ranked the drawings according to the level of personal adjustment which they felt was reflected in the drawings. They also ranked the drawings according to degree of artistic excellence. Commercial artists also ranked the drawings for artistic merit. The results were as follows.

1. Extremely reliable evaluations were achieved by all judging groups.

2. The clinical psychologists utilized the artistic merits of the human figure drawings to a large extent in their adjustment evaluations of them.

3. The psychologists judged the drawings in much the same manner, whether consciously judging according to art or adjustment.

4. Neither art nor adjustment ratings by artists or psychologists show any consistently significant relationship with the dichotomy psychiatric patient vs. nonpsychiatric subject. The art values sometimes depress, sometimes inflate, the relationship.

5. As judged by the "average" clinical psychologist today, human figure drawings executed by persons of average or above-average intelligence seem to indicate art achievement but do not seem to show any consistent relationship to level of personal adjustment.

Received February 23, 1953.

References

1. Albee, G. W., & Hamlin, R. An investigation of the reliability and validity of judgments inferred from drawings. *J. clin. Psychol.*, 1949, 5, 389-392.
2. Edwards, A. H. *Experimental design in psychological research.* New York: Rinehart, 1950.
3. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
4. Roe, Anne. The use of clinical diagnostic techniques in research with normals. In M. L. Reymert (Ed.), *Feelings and emotions.* New York: McGraw-Hill, 1950. Pp. 336-342.

Journal of Consulting Psychology
Vol. 17, No. 6, 1953

The Examiner's Influence on the Rorschach Protocol [1]

Robert G. Gibby
Veterans Administration Mental Hygiene Clinic, Detroit, Mich.

Daniel R. Miller, and Edward L. Walker
University of Michigan

Psychotherapists are generally agreed on the principle that the therapist is personally involved in the relationship as much as the patient. According to Jung [4], both are mutually influenced in the treatment. Sullivan [9] regards the therapist as a "participant observer." In Rogers' opinion, ". . . to treat another person as a person is to open oneself to change through the influence of the relationship" [7, p. 175]. From this conception of therapeutic interaction, Freud [3] and Oberndorf [6] deduce that even competent therapists may occasionally be unsuccessful because of certain aspects of their physical appearances, their personalities, or their blind spots.

Like the therapist, the administrator of the Rorschach test may also tend to elicit certain reactions from his subjects. In that event he should obtain protocols that are different from those given by subjects of examiners with different characteristics. The present paper is concerned with this hypothesis of a significant over-all difference in protocols obtained by various examiners even when they are testing similar groups of subjects. This has already been demonstrated by Cleveland [2] and Sanders [8] for relatively inexperienced examiners who tested college sophomores under the highly controlled and presumably artificial conditions of an experiment. The purpose of the present study was to test this finding with experienced examiners who administered the tests under the standard conditions of the clinic to equated groups of psychiatric patients.

Subjects

Patients. Cases were selected from among the most recent patients tested by each of 12 examiners at the Mental Hygiene Clinic of the Veterans Administration Regional Office, Detroit, Michigan. The assignment of cases to the clinical psychologists at the Detroit Clinic is determined by the days of the week when the patients are first admitted. Since there is no policy of scheduling any particular type of complaint on a given day, each psychologist had an equal chance of being assigned patients in all the diagnostic categories. A tabulation of cases of the different examiners revealed no significant differences in the proportions of various pathologies. The diagnoses of the final sample range from simple neurosis to prepsychotic condition. To facilitate the interpretation of results all subjects who were selected were male, white, between the ages of 25 and 32, not disoriented, and their symptoms were functional rather than organic.

Psychologists. Three standards were used in the selection of examiners. The first was that they have a minimum of two years of experience at the clinic. This helped to insure a common theoretical frame of reference and the use of similar techniques. They all administered the Rorschach test according to Beck's [1] instructions. Second, clinicians were selected only if they had tested at least 20 cases who met the criteria for the choice of subjects. Finally, an attempt was made to obtain divergent personalities on the basis of ratings of the psychological staff by two of the writers. There were nine males and three females.

Method

All records were first coded so that they contained no obvious clue to the identities of the examiners or subjects. One of the authors then scored all the protocols by Beck's method. To provide for the calculation of reliabilities, a randomly selected sample of records was also scored by a second person.

---

[1] From VA Regional Office, Detroit, Michigan.

The data were analyzed in two ways. First, the raw or absolute scores of the determinants *R, F, M, FC, CF, C, FY, YF,* and *V* were used in analyses of variance to determine whether examiners differed significantly from each other in the values they produced in their protocols. The *F*'s for each of these determinants are contained in Table 1. It was originally planned to include *Y,* but this scoring category was not used because of the paucity of responses in this category.

Second, in order to control for unequal numbers of responses, separate analyses were made of the proportions of scoring categories in relation to lengths of protocols. Since all data in the percentage column are ratios of number of responses, there can obviously be no entry under *R*. Because percentages are not normally distributed, they were converted to arc-sine units. Then, as in the cases of the raw scores, analyses were made of the variances among examiners of arc-sines of the eight determinants. The *F*'s for these analyses are also contained in Table 1.

Results

The interjudge reliabilities have a mean of .91 and range between .66 and 1.00. Only three of the nine are below .90. The low correlation of .66 is attributable in part to the small number of *YF* responses, only 86 in 240 records.

Table 1 lists variances among examiners for absolute and percentage scores. Of the nine absolute scores, variances among examiners for three, *F*, *C*, and *FY*, are significant at the 5 per cent level. This is approximately seven times the number of significant variances that might occur by chance. Variances among examiners on three of the eight percentage scores, *F*, *C*, and *YF*, are significant at the 5 per cent level. This is approximately eight times the number that would be expected by chance. The results thus indicate that there are significant over-all differences in the determinants obtained by various examiners from comparable groups of subjects in an outpatient mental hygiene clinic.

Some indications of the magnitudes of differences among examiners are afforded by the ranges of means of the significant determinants. In the case of *F*, there is a considerable range of means, from 10.0 to 20.8. If *F* is given its conventional interpretation as an indicator of control, the possibility arises that when testing members of a homogeneous population some examiners will elicit protocols which indicate overconstriction while the records obtained by others will indicate weak control. In view of the range of means for *FY*, 1.0 to 3.4, it is also probable that certain examiners in this clinic tend to obtain excessively dysphoric records while others with comparable patients may seldom elicit such reactions.

It is more difficult to assay the significance of the results for *C* and *YF*, the determinants with smaller ranges. Means of *C* obtained by the different examiners vary between zero and .55; means of *YF* range between .05 and .70. The absolute magnitudes of these ranges are sufficiently small to suggest that the significant variances may be a function of the sample of patients. In the case of *YF* this is made more probable by the previously mentioned small number of responses in that category. Thus, despite the significant variances among examiners for *C* and *YF*, confidence in these results must depend upon the repetition of the study in the same clinic.

Assuming the results are stable, the magnitudes of variances should be judged in terms of current standards of interpretation. In some instances, particularly those involving patho-

Table 1

Differences in Rorschach Determinants of Examiners

Scoring symbol	Absolute score *F*	Percentage score *F*
R	1.39	
F	1.83*	2.00**
M	1.30	.58
FC	1.15	1.42
CF	.60	.06
C	2.06*	2.37
FY	2.56**	1.29
YF	1.40	1.91*
V	1.45	1.65

* Significant at the 5 per cent level or less.
** Significant at the 1 per cent level or less.

gnomonic signs, the presence of certain determinants even in small amounts may affect the interpretation quite radically. This is particularly true of pure color, so that the narrow range of means for *C* may still point to a significant source of error.

Discussion

To what extent can these results be applied to the records obtained in different clinics? The answer to this question is determined in great part by theoretical orientation. In the opinions of the writers, the subject's responses indicate how he organizes the drives aroused in the course of his interaction with a particular examiner in a specific situation. From this premise it may be deduced that the administration of the test by different examiners in the same setting or the same examiner in different settings might produce different protocols. This deduction creates considerable difficulties for the design of corroborative studies and the establishment of norms. However, it also suggests fruitful hypotheses.

If this investigation were repeated in another installation, even identical results could not be viewed as validating those of the present study unless conditions were identical. If the conditions were not identical, significant differences among examiners for more determinants might reflect more disturbed or less experienced examiners. A complete lack of significant variances might result from marked similarities in the interpersonal habits of the examiners, or from the fact that they were experienced. Similarly, the data obtained in connection with variations in setting or in type of patient could not be compared with those of the present sample.

Strictly speaking, then, the present results can apply only to a situation homogeneous with that of the present clinic. This principle should also hold for the establishment of norms. A partial recognition of this fact is afforded by the frequently repeated statement that Rorschach norms can be used only within wide limits. In part, clinicians claim, the absolute magnitudes of the determinants take on different meanings depending upon the contexts or gestalten in which they occur. Moreover, it is the hope of each examiner that, with sufficient experience, he can learn to establish a group of subjective corrective factors that can cancel such sources of error as his idiosyncratic effect on the patient's responses or unusual variations in the settings or patient's motivation. Such a faith, however, is based on the assumption, one with which the writers are inclined to agree, that once the sources of error are eliminated, there are absolute norms which are comparable in different situations.

Rather than stress lack of generality, then, it would be more fruitful to inquire into the conditions under which results can and cannot be generalized. Some light may be thrown on this problem by the basic dimensions that Miller [5] attributes to the interpersonal relationship between examiner and subject. The subject's character structure constitutes one group of variables. Within this group, the degree of pathology of the symptoms may bear upon the degree of variance among examiners. For example, experience has taught most clinicians that, with certain exceptions, the greater the pathology, the more inflexible are the responses of the patient. In other words, regardless of the examiner's characteristics, the schizophrenic or compulsive may tend to order indiscriminately their relationships with him by means of the same defenses with which they structure most of their experiences. Normal subjects, on the other hand, should be more sensitive to the examiner's physical and psychological characteristics. Other things being equal, then, the number of significant variances among examiners should decrease with increasing pathology. This hypothesis should not be difficult to investigate. The Rorschach test might be administered by the same group of examiners in the same setting to three groups of patients, normal, neurotic, and psychotic. Normal patients might be selected from medical clinics where there could be some insurance that their symptoms were not psychogenic.

A second variable that may bear upon comparability of data is the relative experience of the clinicians. If three examining groups, each with differing amounts of experience, were used in the experiment with the three groups of patients, the design could utilize the analysis of variance technique. Not only would this require fewer examiners than other methods, but analyses could be made of each of the variables separately, of their interaction, and of combinations. The value of the latter is illustrated by the prediction that there are more significant variances among inexperienced examiners testing normal subjects than among very experienced clinicians with psychotic patients. In the latter case, significant variances may tend to disappear, thus enabling the clinicians to share norms. Comparable studies of additional dimensions of the interpersonal situations should contribute significantly to the identification and ultimate measurement of the primary sources of error on the Rorschach test.

Summary

This study is concerned with differences in Rorschach protocols obtained by various examiners from comparable groups of patients in a mental hygiene clinic. Factors controlled in the matching of groups were sex, race, age, orientation, and organicity.

Of nine absolute scores of determinants, three, F, C, and FY, are significant at the 5 per cent level. Three of the eight percentage scores, *F*, *C*, and *YF*, have 5 per cent probabilities of occurrence by chance. The significant results are respectively seven and eight times the number that would be expected by chance. In short, the examiners differ significantly from each other in the determinants they elicit from comparable subjects.

A theoretical frame of reference is proposed for the interpretation of results and the planning of further research.

Received February 25, 1953.

References

1. Beck, S. J. *Rorschach's test.* Vol. I. *Basic processes.* New York: Grune & Stratton, 1944.
2. Cleveland, S. E. The relationship between examiner anxiety and subjects' Rorschach scores. Unpublished doctor's dissertation, Univer. of Michigan, 1950.
3. Freud, S. Recommendations for physicians on the psychoanalytic method of treatment. In *Collected papers.* Vol. II. London: Hogarth Press, 1946. Pp. 323-333.
4. Jacobi, J. *The psychology of Jung.* New Haven: Yale Univer. Press, 1943.
5. Miller, D. R. Prediction of behavior by means of the Rorschach test. *J. abnorm. soc. Psychol.*, 1953, **48**, 367-375.
6. Oberndorf, C. P. Unsatisfactory results of psychoanalytic therapy. *Psychoanal. Quart.*, 1950, **19**, 393-407.
7. Rogers, C. R. Where are we going in psychology? *J. consult. Psychol.*, 1951, **15**, 171-177.
8. Sanders, R. The relationship between examiner hostility and subjects' Rorschach scores. Unpublished doctor's dissertation, Univer. of Michigan, 1950.
9. Sullivan, H. S. Conceptions of modern psychiatry. *Psychiatry*, 1940, **3**, 1-117.

Journal of Consulting Psychology
Vol. 17, No. 6, 1953

Personality Factors in Seizure States with Reference to the Rosenzweig Triadic Hypothesis

Francis M. Canter
Walter Reed Army Hospital[1]

There has been a great deal of controversy concerning the role of personality factors in the production and aggravation of epileptic phenomena. Some investigators have tended to regard all seizures as expressions of emotional conflict, whereas others have differentiated between epilepsy as an organic, probably inherited, illness and hysterical seizures in which psychopathology is the predominant feature. Many workers, of course, who have written on this subject make no definite distinctions between organically based seizures and hysterical attacks. Instead, epilepsy and hysteria are seen as opposite poles of a continuum where organic or constitutional factors and psychological maladjustments operate together in an intricate fashion.

The psychoanalytically oriented writers have most consistently emphasized the basic psychological disorder which eventuates in epileptic symptomatology. While they do not deny that organic predispositions may exist, they regard the evidence from electroencephalography, drug treatment, and surgery as of little consequence for a real understanding of the attacks. Fenichel [4, pp. 265-267], Jelliffe [6], Mittelmann [9], Clark [1], and others have pointed to repressed hostility, extreme narcissism, and the regressive nature of the seizure as indicative of its basically psychological origin.

Lennox [7] is one of the main protagonists of the point of view that epilepsy is essentially a neurological disease and should be clearly differentiated from the psychogenic seizure episodes, which may be considered hysterical. Those who hold this viewpoint do not disregard the role of psychological stresses in the aggravation or even precipitation of the convulsive syndrome, but they regard these stresses as no more important than any other stress, physical or environmental, with which the organically weakened organism cannot cope. Consequently, they believe that epilepsy is principally a pathophysiological problem, and that in cases where no brain dysfunction is demonstrable, the term hysteria may be applied. Both epileptic and "psychogenic" seizures may occur, of course, in the same individual, and these cases will present special diagnostic and treatment problems. However, in a recent restatement of his position, Lennox [8] has estimated that 95% of epileptic patients in his experience had no hysterical seizures, alone or in conjunction with the epileptic attacks.

There is increasing evidence that any individual may have convulsive symptoms under the right conditions. Some investigators, e.g., Gastaut [5], have recently attempted to discover the relative "convulsive thresholds" of varying subjects such as idiopathic epileptics, schizophrenics, and hysterics. From this standpoint, the designation of "epilepsy" or "hysteria" becomes an arbitrary one, depending on the level at which it is concluded that the organic weakness is a more deciding factor than the psychological stress which may be involved. Presumably, an individual with a very low convulsive threshold would succumb to minor stresses, whether psychological or physiological, whereas an individual with a high threshold would bear up under all but quite severe stress. Nevertheless, the differential distinction between epileptic and psychogenic seizures continues to be used as a practical prerequisite for therapy.

The present study is an attempt to outline certain personality patterns in two groups of seizure patients. A group of patients diagnosed as idiopathic epileptic will be compared with a group of patients whose seizures are attributed to emotional factors and who are considered hysterical. It is expected that the patients with hysterical seizures will, as a group, show evidence of greater psychological disturbance than the group of epileptic patients. If this hypothesis is sustained, it will provide significant support for the concept that the psychological involvement need not be as pronounced for patients with an organic weakness.

---

[1] The present paper is a condensation of part of a doctoral dissertation submitted to the Graduate Board of Washington University. The author wishes to express his gratitude to the sponsor of the dissertation, Dr. Saul Rosenzweig, for his encouragement and assistance. The author also gratefully acknowledges the many helpful suggestions of Dr. Wilse Webb. The research could not have been accomplished without the constant cooperation and interest of Col. Donald Peterson, MC, USA, and Dr. Earl Swartzlander of the Neuropsychiatric Service, Fitzsimons Army Hospital.

With respect to the personality variables on which the two groups might be expected to differ, the Rosenzweig triadic hypothesis was considered applicable. This hypothesis postulates that the availability of repression as a defense mechanism, the tendency to react to frustration in an impunitive manner, and relative ease of hypnotizability will be found in positive association with each other. The hypothesis is based on observation of the appearance of these factors in hysteria, and on experimental studies [11, 15].

With the hypothesis as a "guidepost," it was postulated that a psychogenic seizure (hysterical) group would (*a*) repress to a greater extent, (*b*) be hypnotizable to a greater extent, and (*c*) react to frustration in an impunitive manner to a greater extent than a group of epileptic patients.[2]

In addition, it was postulated that the psychogenic seizure group would, on projective tests, give more responses indicative of anxious and hostile tensions, which might be assumed to stem from the dynamic needs and conflicts which are denied expression in more overt behavior.

Selection of Subjects and Procedure

Subjects for the study were obtained from the Neuropsychiatric Service of Fitzsimons Army Hospital. All male, military patients admitted to the neurological section with primary symptoms of seizures were subjected to an elaborate and intensive examination with the purpose of distinguishing between those patients whose seizures result from organic dysfunction and those whose seizures are symptoms of psychological maladjustment. The ultimate differential diagnosis was formulated mainly on the basis of neurological and electroencephalographic data. An additional criterion for differential diagnosis was the ability of the subject to recall, under hypnosis, the events of the height of the seizure, for which there is, presumably, true amnesia in the genuinely epileptic patient. A detailed description of the procedure leading to this differential diagnosis has been given by Peterson, Sumner, and Jones [10].

In each case, the subject was seen first for hypnotic sessions by a military psychiatric resident and then by the author for the psychological experimentation. Neither the author nor the hypnotist, at the time of examination, knew anything of the results of any previous tests or examinations upon which the diagnosis might depend, nor the results of the other's examination. The experimental variables, other than recall under hypnosis, did not enter into the final diagnosis.

The two experimental groups for this study were as follows:

1. Forty-four patients diagnosed as having idiopathic epilepsy. This group will be termed the epileptic group.
2. Forty-five patients diagnosed as neurotic or character disorder, whose seizures are considered symptoms of psychological maladjustment. This group will be termed the psycholeptic group.

Since the subjects were tested as they came, without indication at that time of what the ultimate diagnosis might be, it was impossible to match the groups in advance on such variables as intelligence, age, etc. However, the two groups as finally constituted do not differ significantly in these respects. In each group, the average IQ (Wechsler-Bellevue Full Scale) is 101, average age is approximately 23, and average education is slightly over 10 years. On military rank and length of military service, the groups are also comparable.

The order of administration of the experimental measures for each subject was as follows: hypnosis, reaction to frustration (Rosenzweig P-F Study), repression measure, Rorschach test, Thematic Apperception Test (11 selected cards), and Wechsler-Bellevue test.

Hypnosis was measured by the number of hypnotic criteria met by the subject during and after induction of the trance. These criteria include drowsiness, anesthesias, catalepsies, amnesias, hallucinations, and regressive phenomena, these states in this order presumably indicating a progression from a relatively light to a relatively deeper hypnosis. There were 12 criteria in the list. The hypnotist also made an over-all judgment as to whether a hypnotic state was induced, and, if so, whether it was a medium trance or a deep, somnambulistic state.

---

[2] The relationships between the variables of the triadic hypothesis, as they are manifested in the seizure groups, will be discussed in a separate paper.

To obtain a measurement of reaction to frustration, the Rosenzweig Picture-Frustration Study was employed. This is a limited projective technique consisting of a series of cartoon-like situations, each of which presents a common frustrating situation. The subject's written responses for each situation are scored as extrapunitive (outward direction of aggression), intropunitive (inward direction of aggression), or impunitive (denial or dismissal of the frustration), in accordance with the scheme of Rosenzweig [14]. Norms for the test are available. Scoring was done in accordance with the criteria and samples provided in the manual.

Repression was measured by means of the technique employed by Rosenzweig in previous studies [12, 15]. Sixteen jigsaw puzzles, with an object pictured on each, were administered to the subject under the guise of an intelligence test. The subject was told that he was expected to do well on all puzzles, and that his score would be evaluated in terms of other individuals' scores. Various devices to increase tension and promote involvement in the task were used, such as constant reference to the stop watch, frowning, and criticism of the subject's performance. On eight of the puzzles the subject was permitted to finish, but on the other eight, he was forced to stop before finishing and was informed that the time limit was up. The subject was allowed to finish the first, second, fourth, sixth, ninth, tenth, twelfth, and fourteenth puzzles in the series, but was not allowed to complete the remainder. After the last puzzle, there was a 15-minute recess, and then the subject was asked to write down the names of all the puzzles that he could remember. It was assumed that, under the foregoing conditions, completed puzzles would be experienced as "successes" and incompleted puzzles as "failures." Repression was considered as shown if the subject recalled more of his successes than his failures.

The Rorschach test and 11 cards of the 1943 edition of the Murray Thematic Apperception Test were administered to each subject and the protocols analyzed for content and thematic indication of anxious and/or hostile preoccupations. The TAT cards were, in order administered, 2, 14, 12M, 6BM, 18M, 13MF, 5, 18GF, 12B, and 9BM. The Rorschach protocols were scored for content indications of anxiety and hostility in the manner described by Elizur [3]. The TAT stories were scored in a similar manner by indicating whether the theme of each story suggested anxiety and/or hostility, or was "neutral" in content. It should be pointed out, in this connection, that the total anxiety-hostility score is considered by the author to be a more meaningful measure than either the anxiety score or hostility score separately, since each score contains elements of the other. The total score, on the other hand, probably is a relatively stable estimate of the intensity or generality of psychological tensions which may be expressed as anxiety on one occasion and hostility on another.

All tests were scored without knowledge of the final diagnosis of the subject. The scoring on the Picture-Frustration, Rorschach, and TAT was independently checked by a psychology graduate student familiar with the scoring used. A check of a sample of approximately one-third of the cases showed correlations with the author's scoring of .94 to .97 for the P-F variables, .95 for the Rorschach total score, and .90 for the total TAT score.

Results and Discussion

The psycholeptic group repressed to a significantly greater degree than the epileptic

Table 1

Comparison of Groups on Experimental Variables

Experimental variables	Epileptic (*N* = 44) Mean	*SD*	Psycholeptic (*N* = 45) Mean	*SD*	*t*
Repression S-F score	–0.068	1.876	1.622	1.911	4.15***
Hypnosis No. criteria	6.14	4.67	4.87	4.18	1.33
Impunitiveness	28.10	10.35	31.25	11.00	1.37
Intropunitiveness	27.21	8.28	30.70	9.43	1.31
Extrapunitiveness	44.02	14.75	37.95	16.91	1.78*
Rorschach Total A-H score	8.16	6.03	11.78	12.08	1.77*
TAT Total A-H score	12.43	4.04	14.82	4.55	2.62**

* Significant at the .10 level.
** Significant at the .05 level.
*** Significant at the .01 level.

group. In the psycholeptic group, 33 subjects recalled a preponderance of successes, 4 subjects recalled a preponderance of failures, and 8 recalled a preponderance of neither. In the epileptic group, 15 subjects recalled a preponderance of successes, 21 recalled more failures, and 8 recalled a preponderance of neither. Comparison of these frequencies results in a χ^2 value of 16.34, which is significant beyond the .01 level of confidence.

In Table 1 the groups are compared on repression in terms of a score consisting of the number of successes recalled minus the number of failures recalled. In terms of amount of repression as well as mere number of cases of occurrence, the psycholeptics repress more than the epileptics. Both experimental groups were further compared with the results from a nonpatient group reported by Rosenzweig [12]. The mean success-minus-failure score of this group was 0.167 with a standard deviation of 1.683. There is no significant difference between the nonpatient, "normal" group and the epileptic group, whereas the psycholeptic group represses more than the nonpatient group, the difference being significant at the .01 level of confidence.

The epileptic and psycholeptic groups did not differ in significant degree with respect to hypnosis (Table 1). In terms of the dichotomization of hypnotized versus nonhypnotized, exactly 25 subjects in each group were hypnotized to a medium or somnambulistic level.

In the interpretation of these results certain factors of possible importance should be noted. When an epileptic subject was hypnotized, he generally met all the criteria up to a given level without gaps or tests which he could not or would not meet. A large proportion of the hypnotizable psycholeptics, however, met the criteria in an erratic fashion, meeting some criteria supposedly indicating deeper hypnosis but not meeting criteria indicating lighter hypnosis. A recent paper by Ehrenreich [2] suggests that this phenomenon may be ascribable to unconscious resistances to those particular activities. If this is the case, it would seem that unconscious resistance to hypnosis was operating to a much greater extent in the psycholeptic group than in the epileptic. In support of this possibility, there are some behavioral and questionnaire data (not reported here) on the present subjects to suggest that the psycholeptics were, in general, more resistive to treatment and examination, possibly because of a realization that their seizures might be gainful to them and/or because of their greater difficulty in adjustment and establishment of interpersonal relations. Of all the measures, hypnosis was probably the most susceptible to interference from situational factors, because of the preconceived attitudes toward it held by many of the subjects and because many apparently feared this seemingly mysterious and unconventional procedure.

In Table 1 is also presented a comparison of the two groups on the variables of impunitiveness, intropunitiveness, and extrapunitiveness. In addition to comparing the groups with each other, each group's scores were compared with the Rosenzweig revised norms [13]. None of the epileptic mean scores differs significantly from the reported norms. However, the psycholeptic group is less extrapunitive than the norm (mean: 45; standard deviation: 13.3). This difference is significant at the .01 level of confidence. Likewise, the psycholeptic group is more impunitive than the norm (mean: 27; standard deviation: 9.45), this difference being significant at the .05 level of confidence. The psycholeptic group shows a tendency to be more intropunitive than the norm (mean: 28; standard deviation: 8.25), though this difference does not reach a level of statistical significance.

It was expected that the psycholeptic group would be more impunitive and less intro- or extrapunitive, but this was not the case. Instead, there is a tendency for the psycholeptics to be more intropunitive and impunitive and less extrapunitive than the epileptics. The division seems to be one of extrapunitiveness versus nonextrapunitiveness, or restraint of outward aggression rather than restraint of all aggression. This may be a significant finding in view of the interpretation of the psychogenic seizure as a turning inward of pent-up hostility.

A factor to be considered here is the ambiguity of the intropunitive response itself, a point previously noted by Rosenzweig and Sarason [15]. In many instances, the intropunitive response on the P-F Study seems to be an exaggerated impunitive type of reaction rather than a reflection of real guilt feelings. Such answers as "I'm very sorry" have become so conventional and stereotyped that with many subjects they appear to mean only an easy way of getting out of the frustrating situation or glossing over the problem, in which actual self-aggression is probably absent. Thus, the intropunitive and impunitive responses may represent the same nonaggressive trait. In this sense, the essence of the original postulation, restraint of hostility, appears to have been substantiated to some degree.

A comparison of the groups with respect to psychic tension, as indicated by the Rorschach and TAT anxiety-hostility scoring, is presented in Table 1. The comparisons are made in terms of the Total A-H score, which, as previously mentioned, is considered a more stable measure of the degree to which psychological tension is expressed in the content and the degree to which neutral or pleasant concepts are abandoned.

There was no difference in mean number of Rorschach responses given by each group, but there was a definite tendency for the psycholeptic group to give more responses with anxious-hostile content. Similarly, the psycholeptics told more anxious-hostile TAT stories and fewer neutral or pleasant stories. In general, then, "psychic tension" appears to be greater in the psycholeptic group.

Table 2

Correlation between Psycholepsy and Experimental Variables

Experimental variables	r_{bis}	p
Repression S-F score	.58	.01
Hypnosis No. criteria	–.03	N.S.
Extrapunitiveness	–.25	.05
Intropunitiveness plus impunitiveness	.22	.10
Rorschach Total A-H score	.25	.05
TAT Total A-H score	.35	.01

In Table 2 the results of the study are presented as biserial correlations between a diagnosis of psycholepsy and increasing scores on the different variables, with psycholepsy-epilepsy as the forced dichotomy. The single most discriminating variable is repression, a finding which suggests that this may be a more basic personality factor than the others in the differentiation of epileptic from psychogenic. The repressive aspect of the seizure is certainly the most immediate and outstanding feature, i.e., the "black-out," amnesia, etc., whereas the factor of repressed hostility is a derived or inferred construct which becomes obvious only after considerable study and interpretation of various behavior patterns.

In spite of the lack of difference between the epileptics and psycholeptics on the hypnosis variable, the latter group presents a pattern of response consistent with what might be expected of a hysterical group. The most definite difference is on repression, which has from the first been linked with hysterical disorders. The evidence for repressing or suppressing aggression, as shown on the Picture-Frustration Study, is consistent with the interpretation of the psychogenic attack as a release or expression of pent-up affects which the subject does not consciously recognize. The finding of greater "psychological tension" in the psycholeptics, as shown by the Rorschach and TAT responses, is consistent with the Picture-Frustration results, if one assumes that the responses on the Rorschach and TAT reflect more unconscious attitudes, whereas the P-F responses are closer to conscious realization of their meaning. The making of stories or the interpretation of inkblots is apt to be farther from the subject's conscious control or censorship of his responses than his answers to the P-F situations, which reflect to some degree a knowledge of, and willingness to abide by, social conventions.

Although there were some possibly invalidating factors with respect to the hypnosis, there remains the possibility that hysterics are actually no better hypnosis subjects than other individuals, except perhaps under optimal conditions. The increased suggestibility of the hysteric, the importance of which is agreed upon by many workers, is only one of the factors contributing to the hypnosis; other factors are the motivation of the subject, his past and present attitudes toward hypnosis, and his relationship to the hypnotist. With the psycholeptic group, the factors of secondary gain which the seizures furnish, repressed aggression toward the examiner, etc., appear to overshadow any intrinsically greater degree of suggestibility. This finding is not unique, for various studies have failed to link hypnosis exclusively with hysteria, probably because of the complexity of the hypnotic phenomena as well as the dubiously hysterical populations upon which the hypnosis was attempted.

The biserial correlations of Table 2, with psycholepsy-epilepsy as the forced dichotomy, are referable to the assumption that so-called psychogenic seizures and epileptic seizures are symptoms of a combination of factors which operate on a continuum from "almost pure" organic "causation" to "almost pure" psychological "causation." The two groups do differ in terms of group averages, but the overlap in the psychological factors is considerable. In addition to the obvious imperfections of the measuring instruments, it is also clear that there is no true dichotomy of psychological involvement, but that the groups probably differ because the subjects are drawn from different points on the continuum. Gastaut's findings, previously mentioned, clearly bring out the increasing importance of the concept of continuous degrees of convulsive susceptibility. The present study, in a gross fashion, approaches the problem from the opposite aspect in its suggestion of psychopathology paralleling the degree of convulsive susceptibility; ordinary or normal psychological adjustment goes with a low convulsive threshold (shown by the abnormal EEG record), and increased psychopathology goes with a higher convulsive threshold (normal EEG record). This interpretation could be put to more precise test by determination of the convulsive threshold in the manner described by Gastaut, and by correlation of the degrees of convulsive susceptibility with degree of psychopathology. The extent to which each factor, organic or psychological, is involved might furnish the basis for determination of treatment. It seems most difficult to attempt a differentiation of personality factors as "cause" of the seizures or as "results," since any personality maladjustment may serve as a source of stress for the organism, and continued stress may even contribute to increased susceptibility, if only through learning and not organic change.

Summary and Conclusions

Two groups of seizure patients were compared for evidence of psychological differences. In one group, termed epileptic, the seizures were diagnosed as symptoms of disturbed brain functioning. In the other group, termed psycholeptic, the seizures were diagnosed as symptoms of psychological maladjustment.

It was hypothesized that the psycholeptic group would (*a*) repress, (*b*) be hypnotizable, (*c*) react to frustration in an impunitive manner, and (*d*) show evidence of psychological tension, to a greater extent than the epileptic group. The hypothesis was sustained for (*a*) and (*d*). The psycholeptic group did not differ from the epileptic group on impunitiveness as hypothesized, but was significantly less extrapunitive. This last finding was interpreted as indicating that instead of suppression of all aggression, only suppression of outwardly directed aggression is involved. No differences between the groups on degree of hypnotizability were demonstrated.

In general, the results indicate that certain patterns of psychological reaction are more prevalent in seizure patients in whom no evidence of disturbed brain function can be found. The personality pattern of the psycholeptic patients is consistent with the theoretical pattern for hysterics. There is also tentative evidence that the psycholeptic group deviates from the normal with respect to the variables of repression and reaction to frustration, whereas the epileptic group does not appear to differ from the normal in these respects.

There is thus some support for the differentiation of seizure patients as organogenic or psychogenic, but the considerable overlap between the groups on the variables studied suggests that the differentiation is not a clear-cut one, and that it applies mainly to the extremes of a psychogenic-organogenic continuum. It is suggested that different convulsive thresholds may be positively related to different degrees of the psychological maladjustment which serves as the stress adequate to precipitate the seizure.

Received February 17, 1953.

References

1. Clark, L. P. What is the psychology of organic epilepsy? *Psychoanal. Rev.*, 1933, **20**, 79-85.
2. Ehrenreich, G. The influence of unconscious factors on hypnotizability: a case report. *Bull. Menninger Clin.*, 1951, **15**, 45-57.
3. Elizur, A. Content analysis of the Rorschach with regard to anxiety and hostility. *J. proj. Tech.*, 1949, **13**, 247-284.
4. Fenichel, O. *Psychoanalytic theory of neurosis.* New York: Norton, 1945.
5. Gastaut, H. Combined photic and metrazol activation of the brain. *EEG clin. Neurophysiol.*, 1950, **2**, 249-261.
6. Jeliffe, S. Dynamic concepts and the epileptic attack. *Amer. J. Psychiat.*, 1935, **92**, 565-574.
7. Lennox, W. *Science and seizures.* New York: Harper, 1941.
8. Lennox, W. Psychiatry, psychology and seizures. *Amer. J. Orthopsychiat.*, 1949, **19**, 432-446.
9. Mittelmann, B. Psychopathology of epilepsy. In P. Hoch, & R. Knight (Eds.), *Epilepsy.* New York: Grune & Stratton, 1947. Pp. 136-148.
10. Peterson, D., Sumner, J., & Jones, G. The role of hypnosis in differentiation of epileptic from convulsive-like seizures. *Amer. J. Psychiat.*, 1950, **107**, 428-432.
11. Rosenzweig, S. The experimental study of repression. In H. A. Murray (Ed.), *Explorations in personality.* New York: Oxford Univer. Press, 1938. Pp. 472-490.
12. Rosenzweig, S. An experimental study of "repression" with special reference to need-persistive and ego-defensive reactions to frustration. *J. exp. Psychol.*, 1943, **32**, 64-74.
13. Rosenzweig, S. Revised norms for the adult form of the Rosenzweig P-F Study. *J. Pers.*, 1950, **18**, 344-346.
14. Rosenzweig, S., Fleming, E., & Clarke, H. Revised scoring manual for the Rosenzweig Picture-Frustration Study. *J. Psychol.*, 1947, **24**, 165-208.
15. Rosenzweig, S., & Sarason, S. An experimental study of the Triadic Hypothesis: reaction to frustration, ego-defense, and hypnotizability. I. Correlational approach. *Charact. & Pers.* 1942, **11**, 1-19.

Journal of Consulting Psychology
Vol. 17, No. 6, 1953

Research Origin and Construction of the I. P. A. T. Junior Personality Quiz

Raymond B. Cattell and Halla Beloff
Laboratory of Personality Assessment and Group Behavior
University of Illinois

Alike in predicting school progress and in child guidance work a need has long been felt for an adequate and comprehensive personality measuring instrument with which to supplement ability tests. Within the adult range, researches on the nature and measurement of whatever stable, independent dimensions of personality can be shown to exist [3] have resulted in practical instruments for the measurement of personality source traits, such as the *16 Personality Factor Questionnaire* (the 16 P. F. test [11]). But it is only recently that the personality factor structure of 10- to 14-year-old children has become sufficiently clarified, in terms of behavior ratings [9], questionnaire responses [10], and objective tests [8], to permit any truly informed test construction.

The present article deals with the development of a children's personality questionnaire, paralleling the adult *16 Personality Factor Questionnaire* [4, 11] and based on the published factorization of children's questionnaire responses. For although the final aim of the researchers in this laboratory has been to develop objective types of tests, it seems desirable to make available immediately to applied psychology a measure of these primary factors. Indeed, since the objective test battery for 12 factors is likely to require about four hours of testing, it seems that there will in fact always be situations where practitioners will be compelled by practical restrictions to use the more convenient questionnaire form.

In considering this last point—motivational questionnaire distortion—one should bear in mind, however, that in the present J.P.Q. as in the adult 16 P.F., the questions are, in general, indirect and are not accepted at their face value as self-assessments. *Q* data in our studies are no longer on the footing of introspective evidence but are treated as strictly behavioral responses, the meaning of which is to be found, not from the overt verbal meaning, but from correlations with observed behavior in real life situations.

No description will be given here of the data which are being related to the questionnaire responses, for they are set out elsewhere [8, 9], but a brief recapitulation will now be made of the research origin of the test, namely, of the factor analysis of the questionnaire responses upon which the present test construction rests.

Synopsis of the Factor Analytic Foundation

Since our intention was to arrive at a questionnaire of nearer 100 than the usual 200 items—on account of children's limited powers of attention—we planned to begin with about 300 items, which could then be subjected to about two-thirds attrition by the selection process. These were chosen (*a*) to cover the whole field of behavior, i.e., not merely neurotic, etc., areas; (*b*) to be of interest to children; (*c*) to correspond in many cases to items in the adult 16 P.F., being thus already known to measure particular factors. Hence, by apportioning about ten such equivalent questions to each adult factor, we hoped to succeed in cross-matching childhood and adult factors by "marker variables" alone. A 295-item questionnaire was ultimately presented to 330 boys and girls between the ages of 10 and 13 (averaging 11.2 years), using dual-choice answers. This set of items was reduced in number by the following selection processes.

1. All items were dropped which children found difficult to read, or of uncertain meaning.
2. All items were dropped in which the yes/no

cut was more eccentric than 70%–30%.

3. All items were dropped which had no significant correlation with any other item.

These eliminations left 103 questions, which were factorized in a single matrix and rotated successfully to simple structure with respect to 14 factors. At this point, with the factors defined, 115 newly constructed items, which also passed the first two of the above criteria, were correlated on the same population in an "extension of the matrix" by Dwyer's method [14], to provide more items for some factors which seemed likely to have less than 12 significantly loaded items each. In this way about 4 or 5 new items were picked up for each factor, giving a total of 218 factored items.

From this matrix of 218 items factored into 15 factors we aimed to build up the required junior personality questionnaire instrument.

Selection of Factors for Inclusion

The decision as to which factors should be set up in the standardized test was made on the following considerations:

1. Does the factor have large variance, i.e., is there a sufficiency of items having significant loading to make a minimum of a dozen good items for the factor measurement? In the end, three of the last five factors in order of variance were dropped for failure in this respect.

2. Does the factor have adequate split-half reliability? To assess the internal consistency of the factors chosen by the simple structure the 218 items were readministered to a fresh population of 76 children. In this calculation of the split-half coefficient for each factor the scoring was made only on items above a loading of 0.20 on the factor concerned—averaging about 14 items to the factor. These values ranged from practically zero to 0.83 and the lowness of some suggested that further work is required to improve the simple structure of the chosen factorization—a task best performed by independent rotation of a completely new experimental study. In the end the following factors were dropped (from the present questionnaire construction) on this second basis (i.e., falling below a consistency $r = 0.41$): 10, 13, and 14. Factors 4, 6, and 7, which also fell below the 0.41 limit, were, however, retained because they were highly satisfactory in regard to other criteria, and we suspected that sampling error in this relatively small group could have underestimated their true consistency.

3. Could four psychologists independently examining the items collected for each factor arrive at common, verbally consistent psychological descriptions—and perhaps even interpretations—of the factors discovered? This is not an attempt to put new wine in old bottles. As far as interpretation is concerned, if the factor unity does not fit pre-existing clinical or literary conceptions of psychological functional unities, such a finding is accepted principally as a reflection on the validity of the latter. But, at the descriptive level, on the other hand, one has increased confidence in the factor if the same, consistent central characteristic is assigned to it by different perusers of the collected loaded items. Factors 12 and 14 were the only ones in which differences of description and conception occurred, and this caused these, when combined with other defects, to be rejected.

4. Is the simple structure convincing in terms of clearness of hyperplane, reaching a level of 60% or more items in the $+0.10$ hyperplane?

The combined application of these criteria at a severe level would cut the questionnaire down to 5 or 6 factors. Among those lost would be some clearly recognizable, known factors in the adult realm, such as I (Sensitive emotionality), Q_3 (Will control), and H (Cyclothymia) as well as some less confidently labeled factors which are nevertheless of large variance in terms of accounting for child behavior. It would seem a mistake to eliminate from a routine test, simply because split-half reliability is not yet high, any factor which could contribute (with refinement of items) to increasing educational prediction or clinical prognosis. For if and when these predictive powers are confirmed through use and external validation of the test, the reliability coefficients can readily be raised by increasing the number of items and by conventional item analysis. And since expense does not permit the publication of two tests—one, for routine practice, containing only the 6 factors of highest clarity and reliability, and another, for further research, of lesser internal consistency but involving all 15 newly adumbrated factors—it seemed best to compromise, by drawing the line at those 11 factors in which all the above criteria are tolerably satisfied simultaneously.

Thus we decided to set up, in the standard questionnaire, factors 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, and 15, leaving the remaining 3, adequately described elsewhere [10], to await further basic research.

Questionnaire Design and the Consolidation of Factors

With the aim of convenience in use the following considerations of design directed the choice and arrangement of items for the above eleven factors in the actual questionnaire.

1. A uniform 12 items per factor was aimed at and proved to be feasible for all factors, since only on 1 or 2 items on two or three factors did this force us slightly below the chosen lower limit of loading of 0.25. (The per item *average* is naturally much higher.) Twelve is not a very large number of items on which to measure a factor, but it is as much as is compatible with having a questionnaire which can be completed in a class period and within the endurance of a ten-year-old child. Higher reliability is best achieved, as we are planning, by constructing a second equivalent scale also with 12 items per factor. Our first step was therefore to examine the rotated matrix [10] and make tests of about the 24 highest-loaded items for each factor—counting on a 50% rejection rate by further considerations.

2. We desired to avoid spurious correlations among the factor measurements and aimed to achieve this (*a*) by never using one item in more than one factor, (*b*) by balancing plus and minus loadings occurring on an irrelevant factor among items chosen for high loadings on the factor concerned. This involves a good deal of art in combination. For example, item 34 has its highest loading, −.39, on Factor 4, but also loadings of −21 on 15, 26 on 7, and 26 on 9. These are substantially "cancelled" by including in Factor 4, (*a*) Item 146 (.37 on 4) with a loading of .28 on 15; (*b*) Item 79 (−38 on 4) with a loading of −23 on 7; and (*c*) Item 203 (33 on 4) with a loading of −24 on 9.

This cancellation—such that no loading on an extraneous factor was repeated with the same sign, i.e., unsuppressed, in an item chosen for the given factor—was achieved on all but Factor 7 (slight excess of −8) and Factor 1 (slight excess of +6, −7, and −8). In these last cases, where the nature of the intruding factor was well known, an attempt was made to eliminate them psychologically by molding the verbal expression of the item so as not to involve the contaminating tendency.

3. Lastly, response position and affirmative-negative tendencies were balanced within each factor. There was initially some predominance of "Yes" answers operating in the direction of the high factor scores. Where the item could not be inverted without damage to its meaning, it was left untouched and an "*a*" or "*b*" response in the same factor was inverted instead, thus balancing on position of responses.

Except for this latter concession, made in three factors, both position (first and second) and affirmative-negative responses were simultaneously balanced in contributing to the factor score. Whenever inversion of the sentence or meaning had to take place, the opportunity was taken to employ the modification to make the item agree better with the meaning of the factor as it had finally come to be understood. Such changes were kept slight, and were carried out only by a person with long experience in this field.

4. Inspection of the 132 surviving items suggested that in a few cases the secondary processes of (2) and (3) above might have resulted in taking items not altogether adequate in primary loadings. Moreover each factor still contained, on an average, two items of its twelve with loadings initially between 0.25 and 0.30, contrary to our hope of setting the lower limit of 0.30 in the final questionnaire. Accordingly it was decided to set up a new experiment to recruit fresh "replacement" items for those inadequate and below 0.30. For this purpose a further experimental "extension" of 128 new questionnaire items was set up, the items being specially created, on the basis of newly gained psychological insights through the factor analysis.

These were administered to 120 eleven- and twelve-year-old boys and girls, along with the questionnaire as it then stood at 132 items, and a second "extension matrix" was calculated showing the correlations of the 11 factors with each and all of these new items. Whenever a new item correlated 0.30 or more with a factor, without correlating significantly with any other factor, it was used to replace one of the lowest 2 or 3 items in the existing factors. In this way 28 improved items were introduced replacing those deemed inadequate.

This final administration gave another opportunity to ensure that no modified or overlooked item could remain that would be difficult of comprehension by ten-year-old children, and in two cases slight simplifications were found possible.

Table 1

J. P. Q. and 16 P. F. Correlations

	16 P. F. Test Factors															
J.P.Q. Factors	*A*	*B*	*C*	*E*	*F*	*G*	*H*	*I*	*L*	*M*	*N*	*O*	Q_1	Q_2	Q_3	Q_4
1								51*			−38		−30			
2				−43			−44		32			68*			−39	59
3			−60				−51		38	38		62			−38	67
4					−30	48*									44	
5						−23						43			−53	49
6					32						36					
7				32			63*									
8		31	39			36			33	−39					38	−47
9										34						
10				36*												
11	35*					−33										
12		35*														

* Highest *r*'s in row and column.

5. The factors were finally arranged with these items in cyclical order in the questionnaire, for ease of scoring on the answer sheet, with a stencil. The items of easiest decision and reading were arranged at the beginning for "warming up."

6. For uniformity with the adult 16 P.F., and because the inclusion of a brief measure of the intelligence factor along with other factors in the latter has already proved highly convenient in many practical situations, it was decided to include a brief intelligence test here also, as a twelfth factor. For this purpose 12 items were taken from Cattell's earlier validation and standardization of verbal intelligence tests [1, 2], being chosen to subtend the range from 10 to 15 years with maximum discrimination. The total number of items in the test is thus raised to 144, which, at a maximum allowance of ¼ minute per item, should still be within the usual length of a class period. A time-of-working distribution for various ages will eventually be worked out.

7. Although the ultimate interpretations and labeling of these factors must await their correlations with all kinds of child behavior and performance, some preliminary indication of their nature could be gained by correlation with the adult 16 P.F. test factors, which already have many associations. Accordingly the J.P.Q. was administered along with the 16 P.F. to 100 16-year-old highschool boys and girls in the U.S.A. and 56 15-year-old secondary school children in Britain. Only the *r*'s above 0.30 are recorded in Table 1.

To approach a cross-identification of factors from the data yet available in Table 1 is not a simple step. For we have to take account first

Table 2

J. P. Q. Factor Intercorrelations

	1	2	3	4	5	6	7	8	9	10	11	12
1												
2	.35											
3	.18	.44										
4	.12	−.10	−.31									
5	−.18	.18	.39	−.37								
6	−.02	−.05	−.22	.24	−.10							
7	−.13	−.29	−.18	−.03	−.19	.25						
8	.17	.01	−.27	.29	−.38	.17	.04					
9	−.25	.18	.26	−.21	.31	−.09	−.05	−.42				
10	−.30	−.17	−.12	−.05	.12	.09	.09	−.09	.07			
11	.17	.06	.15	−.14	.09	−.04	.04	−.23	−.09	.26		
12	.02	.11	−.09	.13	−.06	.01	−.12	.08	−.10	−.08	−.04	

of the corrections for attenuation justified by the marked brevity of the factor measures; second of the fact that any factor measure is at present still contaminated with others; and thirdly that even the pure factor measures themselves are slightly correlated. The combined effect of the second and third of these can be appraised from Table 2, representing the correlations among the 12 factors of the J.P.Q., as measured by the final test form upon a new population of 100 twelve-year-olds.

The pattern of intercorrelations has a slight resemblance to the pure interfactor correlations obtainable from the matrix [10] but in the main the intercorrelations are larger and evidently due to some items in each factor still being appreciably loaded in other factors too. Further research will be directed to purifying factors by replacing these items. Meanwhile we know that the highest intercorrelation is 0.44 and some 46 of the 66 interrelations are below 0.20, which is below significance (1%) for this size of sample. Parenthetically it is noticeable that the significant correlations—0.44 between Factors 3 and 2, and 0.39 between 3 and 5—are confirmed by these factors having similar profiles in their correlations with 16 P.F. factors in Table 1.

To seek cross-identification we may first make a rough correction for attenuation effects[1] by taking the *r*'s in Table 1 that are simultaneously highest in their row and column, and marking this with an asterisk. This gives a match of J.P.Q. Factor 1 with 16 P.F. Factor *I* and, similarly, 2 with *O*, 3 with *C* and Q_4 combined, 4 with *G*, 5 with Q_3, 7 with *H*, 10 with *F*, 11 with *A*, and 12 with *B*. Thus in 9 of the 12 new factors there is no alternate 16 P.F. factor which would compete with these interpretations, while in the remaining 3 the identification must be considered indeterminate, 6 possibly matching *N*, 8 possibly *L*, and 9 possibly *M*. The one unsatisfactory conclusion above is in having to match 4 with a combination of *C* and Q_4 (though its resemblance in meaning to both was actually noted, before correlation, in the former article [10]). *C* and Q_4 are alike in being labeled the neuroticism factors in the adult-population studies and in having somewhat similar effects on social situations, e.g., in most potently reducing an individual's chances of being elected a leader [12].

The complexity of the manifest correlations in Table 1 beyond what would be expected from a simple one-to-one match, could arise from (*a*) the above brevity-attenuation effects; (*b*) the factor structure in preadolescent children being actually different from that of adults; (*c*) the intercorrelations of the equivalent child and adult factors being different; (*d*) the measures being contaminated in one or both scales; and (*e*) the two scales, adequate in their proper age ranges, being distorted in the intermediate age where we try to make them meet.

Only further experience and research can clarify these issues. As indicated, a study is in progress to give an independent contribution to the J.P.Q. rotation, admittedly not perfect, and to replace contaminating items, which should truly align the J.P.Q. and 16 P.F. factors now matched. On the other hand, with respect to the difficulty in J.P.Q. Factor 3 (with two marked *r*'s in its row in Table 1) it is psychologically conceivable that two patterns of adult neuroticism develop out of a single "protoneuroticism" factor in children in the latent period, which factor is perhaps derivable from the adult data as a second-order factor. Parenthetically, the fact that in two cases the *r*'s accepted for matching fall as low as 0.35 should not worry us when the whole configuration of correlations is sound. Actually, one of these two lowest is between intelligence in the J.P.Q. (Factor 12) and intelligence in the 16 P.F. (Factor *B*), each constructed of a dozen standard test items! If test brevity, and application of both tests to the limits of their ranges, can do this to such a well-known, tried test factor as general intelligence, then the *r*'s averaging 0.6 among the newer personality factors are quite compatible with a matching *r* of unity between the two factor measures of these personality dimensions, with lengthened scales from the present tests.

---

[1] A more precise correction for attenuation can be obtained from the reliability coefficients reported earlier [10], which (Spearman-Brown corrected) are, for the factors in order: (I) .71, (II) .83, (III) .64, (IV) .36, (V) .63, (VI) .26, (VII) .38, (VIII) .39, (IX) .54, (X) .44, (XI) .56, (XII) .54 on an eleven-year-old population.

Summary

1. A multiple-factor, simple-structure factor analysis has been carried out (with extension devices) on 330 ten- to fourteen-year-old boys and girls on 423 groomed questionnaire items carefully chosen to represent the important dimensions of behavior in the total personality sphere.

2. By dropping 291 items unsatisfactory for eccentricity of cut, ease of comprehension, or goodness of loading, as well as items for 4 factors of defective variance or stability, a questionnaire of 144 items measuring 12 factors by 12 items each has been constructed (adding 12 items for intelligence). By correlation with the adult 16 P.F. these factors are revealed to correspond with such well-known factors as general intelligence (B), general neuroticism (C and Q_4), emotional sensitivity (I), adventurous cyclothymia (H), anxiety level (O), will control (Q_3), super-ego strength (G), etc., though slight adjustments are still required for proof of full alignment.

3. The usual test-presentation requirements of balancing "yes" and "no" answers within each factor, balancing suppressor loadings for irrelevant factors, eliminating undue vocabulary and reading-time difficulties, arranging items in simple cyclical patterns for scoring ease, and presenting simple instructions and examples for test and machine score-form use, have been worked out. The final J.P.Q. test (Junior Personality Quiz [12]) has 144 items; is suitable for children of 10-16 years; and permits completion in 30 minutes by 90% of children (but may require up to an hour for a small minority).

The external or social validity of this internally validated test remains to be investigated by use of the test in schools and clinics. However, as Cronbach [13] and others have pointed out in terms of information theory, the fact that the test deals with so many demonstratedly independent dimensions of behavior argues for the whole giving better predictions than any single, long, and reliable scale, because each factor brings new information. Thus the use of only 12 items for the intelligence test and of the remaining 132 for eleven further dimensions should increase clinical and educational predictions decidedly above that obtainable from using this number of items (and the rest of the available testing time) in increasingly meticulous measurement of, say, intelligence alone, or in some personality test not demonstrated to deal with more than one or two factors.

A temporary standardization has been issued with the test and a further age and sex standardization on 2000 cases is in progress. Social extension of the meaning of the factors is also at hand in studies in press [6, 8, 11] relating the questionnaire factors to behavior rating and objective test factors obtained for the same population.[2]

---

[2]The writers wish to express their thanks to the teachers and psychologists of Mooseheart and of the Decatur School System, for providing populations for experimentation.

Received February 16, 1953.

References

1. Cattell, R. B. Occupational norms of intelligence and the standardization of an adult intelligence test. *Brit. J. Psychol.*, 1934, **25**, 1-28.
2. Cattell, R. B. Standardization of two intelligence tests for children. *Brit. J. Psychol.*, 1936, **26**, 263–272.
3. Cattell, R. B. *The description and measurement of personality*. Yonkers, N. Y.: World Book Co., 1946.
4. Cattell, R. B. The main personality factors in questionnaire, self-estimate material. *J. soc. Psychol.*, 1950, **31**, 3-38.
5. Cattell, R. B. Research designs in psychological genetics with special reference to the multiple variance method. *Amer. J. Hum. Genet.*, 1953, **5**, 76-93.
6. Cattell, R. B., & Beloff, J. The inter-relations in children of personality factors found in rating, questionnaire and objective test media. (In preparation)
7. Cattell, R. B., Flint, D., & Beloff, J. *The Junior Personality Quiz; handbook and tests.* Champaign, Ill. (1608 Coronado Dr.): Institute for Personality and Ability Testing, 1953.
8. Cattell, R. B., & Gruen, W. Measurement of the primary personality factors in children by means of objective tests. *J. Pers.*, in press.
9. Cattell, R. B., & Gruen, W. The personality factor structure of 11 year old children in terms of behavior rating data. *J. clin. Psychol.*, in press.
10. Cattell, R. B., & Gruen, W. The primary personality factors in the questionnaire medium for 12-14 year old children. *Educ. psychol. Measmt*, in press.

11. Cattell, R. B., Saunders, D. R., & Stice, G. *The Sixteen Personality Factor Questionnaire.* Champaign, Ill : Institute for Personality and Ability Testing, 1950.
12. Cattell, R. B., & Stice, G. *The behavior of small groups.* Off. Nav. Res. Rep. Washington, D. C.: Dept. of the Navy, 1953.
13. Cronbach, L. A general psychometric theory based on information measurement. Mimeogr. publication, Bureau Res. and Serv., Coll. Educ., Univer. of Ill., 1953.
14. Dwyer, P. S. The determination of the factor loading of a given test from the known factor loadings of other tests. *Psychometrika,* 1937, 2, 173–178.

Journal of Consulting Psychology
Vol. 17, No. 6, 1953

Manifest Anxiety Changes Concomitant with Client-Centered Therapy [1]

James J. Gallagher
Michigan State College

Any study dealing with anxiety must show its recognition of the two-headed aspects of the concept and make clear which of the many possible definitions is being considered in the study. May [7] has pointed out that anxiety can be considered either as a theoretical motivating factor, present by inference rather than by direct observation, or as a descriptive term for a syndrome of physiological and psychological tensions. It is primarily the latter that will be considered in the present study, and the term *anxiety stress* will be used to differentiate it from the term *anxiety* used as a construct in motivation theory.

Most of the newer texts, together with the majority of recent literature in clinical psychology, have tended to emphasize the importance of dealing with dynamic personality qualities such as anxiety, guilt, ego-strength, etc., for personality predictive purposes. Correspondingly, less and less interest is being placed on standard diagnostic or nosological terms such as psychopath, manic, anxiety neurotic, etc. Projective techniques have been especially valuable in focusing attention on these dynamic qualities, yet few of these techniques have been developed to measure a quality such as anxiety stress directly. Most of the efforts of the projectivists have been bent toward discovering the underlying qualities of the personality, and the anxiety discussed is almost always anxiety by inference.

Wechsler and Hartogs [11] have developed and used short tests of motor skills, level-of-aspiration tasks, Draw-a-Person tests, etc. to differentiate between anxiety neurotics, conversion hysterics, obsessive-compulsives, involutional depressives, and incipient schizophrenics. Although many of these measures show some real promise once the quantification problem is solved, they are still in the experimental stage.

Much more popular clinically is the use of techniques already designed for other purposes to measure anxiety stress. Rashkis and Welsh [8] developed a means of detecting anxiety by use of the Wechsler-Bellevue scale; Elizur [1] has produced a content-scoring technique of scoring the Rorschach for anxiety and hostility; and a recent study [3] seems to confirm the promise of such an adaptation.

Probably the most popular technique used to obtain a measure of anxiety stress is the MMPI. Taylor [9] constructed an anxiety scale by having judges choose from the MMPI those items they considered as indicative of manifest anxiety. An item analysis of these items pared the number of items on the final scale to 50. Winne [13] constructed a neuroticism scale with items derived from the so-called neurotic triad which differentiated at the 1% level of confidence between 140 neurotics in a VA mental hygiene clinic and a matched group of 140 normals. A recent study [5] measuring the relationship between the Winne and Taylor scales found them to have a correlation approaching unity.

Other attempts [4, 6] have been made to distinguish the amount of anxiety in a patient by using combinations of the traditional MMPI scales instead of attempting to develop a new anxiety scale. Welsh [12] has approached the problem of profile analysis of anxiety with the most sophistication and has developed two measures, an Anxiety Index and an Internalization Ratio, which seem to differentiate effectively between a wide variety of clinical and normal populations.

Problem. The purpose of the present study is to determine if there have been changes in the anxiety-stress level, as measured by MMPI scales, of a group of college students, concomitant with the application of client-centered therapy.

It is rather difficult to get even tentative agreement among various schools of thought on the goals of therapy; however, practically all disciplines mention the diminution of anxiety stress as an important subgoal. Certainly the removal of anxiety without a subsequent change in the ability of the individual to adapt is of limited value. It is equally true that without anxiety-stress reduction there is little real chance that a patient will be able to adapt himself more efficiently. It was felt by the writer that a constructive approach toward the merging of objectives of various schools of therapy would be obtained by measuring the effectiveness of each school in attaining the measurable subgoals that they claim to achieve by their methods.

---

[1] Data for this study were obtained by the Pennsylvania State College research group, whose original studies were published in *Group Report of a Program of Research in Psychotherapy,* Pennsylvania State Coll., Psychotherapy Research Group, 1953.

Hypotheses.

1. College students undergoing client-centered therapy will show significant decrease in anxiety stress following such therapy.

2. There should be a relationship between reduction in anxiety and therapy success. Therefore, a relatively high correlation would be expected between reduction in anxiety-stress scores and success in therapy.

Method

Subjects. The 42 subjects involved in this study were all students of the Pennsylvania State College who came to the Psychological Clinic between September of 1949 and August of 1950, either by request of various agencies or of their own volition, to obtain aid in their personal adjustment. All these subjects were counseled by advanced graduate students in clinical psychology who had been trained in client-centered therapy. No attempt was made to diagnose these cases prior to therapy except that persons judged as prepsychotic or psychotic on the basis of an intake interview were transferred to the staff psychiatrist. The median number of therapy interviews was between five and six, although some cases were seen for a much longer period of time.

Procedure. A battery of tests consisting of the Rorschach, MMPI, and the Mooney Problem Check List was given to each of the subjects before the beginning of the therapy interviews or, at the latest, before the second therapy interview. The posttherapy tests were given after an agreement was reached by the therapist and client that therapy contacts should end.

The measures used in the present study were:

1. The Taylor Anxiety Scale. This scale was not used in its entirety since many of the present subjects finished only 366 of the total MMPI items. This necessitated using only 34 out of the total of 50 items.

2. The Winne Neuroticism Scale. This scale was included because of its high imputed relationship with the Taylor scale, and to find out if "neuroticism" in this case was synonymous with anxiety stress.

3. The Welsh Anxiety Index. This scale was in reality a combination of the standard scales already in use on the MMPI. This score was obtained by the formula:

$$AI = \frac{Hs + D + Hy}{3} + (D + Pt) - (Hs + Hy).$$

Since this scale requires no scoring other than the original scales of the MMPI, it could obviate the use of the Taylor and Winne scales if its predictive ability were the same or better than those scales.

4. The Welsh Internalization Ratio. This was obtained by the following formula:

$$IR = \frac{Hs + D + Pt}{Hy + Pd + Ma}.$$

This measure was included to see whether the Internalization Ratio would change in the expected direction, toward externalization. As can be seen by the formula, the higher the score the greater the internalization.[2]

The standard t test of significance was used to determine the significance of the differences between the pretherapy and posttherapy measures. The differences on each of the preceding scales were correlated by the Pearson product-moment r method with a multiple criterion for success in psychotherapy to see which of the measures predicted best to that criterion. Correlations were also calculated between the individual measures which went to make up the multiple criterion and pretherapy to posttherapy differences on the anxiety-stress scales.

The therapy criteria measures. A multiple criterion for evaluating client-centered psychotherapy with college students was developed by Tucker [10] using the cases in the present group. This multiple criterion consisted of four measures:

---

[2] The scores obtained by this population on the traditional scales of the MMPI were given in a previous article by the present writer [2].

1. A 29-item therapy rating scale scored by the therapist following the end of contact with the client.
2. The same therapy rating scale scored by judges who read at least 60% of the transcribed interviews, including the first and last interviews.
3. A 14-item client rating scale which the client filled out at the time he took the posttherapy test battery.
4. The *P/N* ratio, which was the ratio of negative to positive feelings in the final interview as compared to the ratio of negative to positive feelings in the first interview.

These four criterion measures were combined into a total criterion score by weighting each measure according to the ratio of its summed intercorrelation with the other three measures. All the rating scales were constructed especially for this study by members of the group research project. For further information concerning the reliability of judges' ratings, the construction of rating scales, etc., the reader is referred to Tucker's original work [10].

Table 1
Differences between Pretherapy and Posttherapy Anxiety-Stress Scores

Measure	*N*	Mean	*SD*
Taylor Anxiety			
Pretherapy	42	17.28	6.18
Posttherapy	42	13.76*	6.13
Winne Neuroticism			
Pretherapy	41	9.85	4.28
Posttherapy	41	7.97*	3.82
Welsh Anxiety Index			
Pretherapy	42	80.60	24.78
Posttherapy	42	70.40*	24.60
Welsh Internalization Ratio			
Pretherapy	42	105.76	15.86
Posttherapy	42	98.41*	15.53

*Difference between pretherapy and posttherapy scores significant at 1% level of confidence.

Results

1. Table 1 shows that all the scales used in the present study reveal a difference in the predicted direction between pretherapy and posttherapy scores that is significant at the 1% level of confidence. Of particular interest is the change in the Welsh Internalization Ratio in which the group shows a definite trend towards externalization following client-centered therapy.

2. Table 2 reveals the correlations between anxiety-stress change scores and the therapy-success criterion measures. As might be expected, the various anxiety-stress measures show a moderate to high agreement among themselves. It is clear that they are measuring somewhat the same quality, but they could hardly be used interchangeably on the basis of the present findings. Of particular interest is the high correlation between the Taylor scale and Winne scale confirming previous findings. One might think that this could be accounted for on the basis of considerable item overlap in the two scales. However, a check revealed that only six of the 34 items on the Taylor scale also appear on the Winne scale.

3. The correlations between anxiety stress changes and the therapy-success measures show considerable variation. The Taylor scale seems to agree with these measures most adequately, although the Winne scale shows a significant agreement with every measure but the client rating scale.

4. The Welsh Anxiety Index and Internalization Ratio did not fare so well, showing relatively low agreement with many of the criteria measures. One is forced to conclude from this that the scales designed specifically to measure neuroticism and anxiety stress are better predictors of therapy success than are combinations of previously used MMPI scales.

Discussion

The results of the present study seem to show rather clearly that there is a decrease in anxiety stress concomitant with client-centered therapy. As was stated previously, this result will have more meaning when placed beside other studies which show behavioral changes and perceptual differences as a result of therapy. Also important would be the same type of comparison done by members of an opposing theoretical and therapeutic orientation.

Table 2
Intercorrelations of Anxiety-Stress Changes and Criteria for Therapy Success

Measure	Winne Neuroticism	Welsh Anxiety Index	Welsh Internalization Ratio	Multiple Criterion	Therapist Rating	Judges Rating	Client Rating	P/N Ratio
Taylor Anxiety	.62	.49	.38	.54	.35	.47	.48	.49
Winne Neuroticism		.30	.58	.54	.51	.45	.21	.42
Welsh Anxiety Index			.71	.28	.08	.28	.42	.21
Welsh Internalization Ratio				.40	.13	.33	.16	.37

Note.—For .05 level of confidence, r = .355; for .01 level of confidence, r = .456.

One may reasonably ask if the high correlations obtained between anxiety scales and therapy criterion measures were not due largely to the therapy measures depending upon observation of this very stress reduction. On the client rating scale and the *P/N* ratio this might be of relatively high consequence; however, on the therapist and judges' rating scales the emphasis in judgments was placed on development of insight, planning behavior, understanding of self, etc. Thus we can say with some confidence that these correlations are a result of concomitant rather than similar behavior.

Summary

An attempt was made to see if there were anxiety-stress changes, as measured by various MMPI anxiety scales, from pretherapy to posttherapy, in 42 college students who underwent client-centered therapy. Comparisons were also made between the change in stress measures from the pretherapy test to the posttherapy test, and the various therapy-success criterion measures.

It was found that, although all four measures showed a significant decrease in stress from pretherapy to posttherapy, two of the measures, the Taylor Anxiety Scale and the Winne Neuroticism Scale, showed the highest amount of agreement with the therapy-success measures.

Received March 30, 1953.

References

1. Elizur, A. Content analysis of the Rorschach with regard to anxiety and hostility. *J. proj. Tech.*, 1949, **13**, 247-284.
2. Gallagher, J. J. MMPI changes concomitant with client-centered therapy. *J. consult. Psychol.*, 1953, **17**, 334-338.
3. Gorlow, L., Zimet, C. N., & Fine, H. J. The validity of anxiety and hostility Rorschach content scores among adolescents. *J. consult. Psychol.*, 1952, **16**, 73-75.
4. Gough, H. G. Diagnostic patterns on the Minnesota Multiphasic Personality Inventory. *J. clin. Psychol.*, 1946, **2**, 23-37.
5. Holtzman, W. H., Calvin, A. D., & Bitterman, M. E. New evidence for the validity of Taylor's Manifest Anxiety Scale. *J. abnorm. soc. Psychol.*, 1952, **47**, 833–854.
6. Hovey, H. B. Somatization and other neurotic reactions and MMPI profiles. *J. clin. Psychol.*, 1949, **5**, 153-157.
7. May, R., *The meaning of anxiety.* New York: Ronald, 1950.
8. Rashkis, H. A., & Welsh, G. S. Detection of anxiety by use of the Wechsler scale. *J. clin. Psychol.*, 1946, **2**, 354-357.
9. Taylor, Janet. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81-92.
10. Tucker, J. E. An investigation in criteria for evaluating nondirective psychotherapy with college students. Unpublished doctor's dissertation, Pennsylvania State Coll., 1950.
11. Wechsler, D., & Hartogs, R. The clinical measure of anxiety. *Psychiat. Quart.*, 1945, **19**, 118-129.
12. Welsh, G. S. An anxiety index and an internalization ratio for the MMPI. *J. consult. Psychol.*, 1952, **16**, 65-72.
13. Winne, J. F. A scale of neuroticism: an adaptation of the Minnesota Multiphasic Personality Inventory. *J. clin. Psychol.*, 1951, **7**, 117-122.

Journal of Consulting Psychology
Vol. 17, No. 6, 1953

A Forced-Choice Form of the Taylor Anxiety Scale[1]

Charles E. Heineman
State University of Iowa[2]

The Taylor Anxiety Scale [26, 27] is a self-rating scale which has shown rather consistent relationships with specific independently defined behavior variables [5, 6, 7, 9, 10, 14, 17, 18, 19, 20, 23, 24, 26, 28, 29, 30]. This scale (referred to hereafter as the *A* scale) prior to its most recent revision [27] consisted of 50 MMPI items selected by judges as being indicative of manifest anxiety according to a definition taken from Cameron [3].

Even though scores on the *A* scale have been shown to possess considerable predictive value in a number of situations, it is possible that they may be influenced by various response sets that are not closely related to anxiety as originally defined by Taylor. One such extraneous variable may be the desire of *S*s to place themselves in a socially favorable light. Thus *S*s may consider not only whether a given statement is actually true of themselves, but also whether their answers will make them look well.

Several findings suggest that *A*-scale scores may be influenced by the variable of social favorability. Among these are the positive skewness of the distribution of scores of the *A*-scale standardization group consisting of 1,971 introductory psychology students who took the *A* scale between 1948 and 1950, and the correlation of $-.74$ between the *A* scale and the *K* scale of the MMPI based on the responses of 281 freshmen at the State University of Iowa. More direct support for the supposition that the expression of anxiety may be considered socially undesirable has been found in an unpublished preliminary investigation.[3]

One method of dealing with the response tendencies associated with considerations of social favorability is to estimate their influence on test scores as is done in the *K* and *L* scales of the MMPI [16]. Another method attempts to reduce the possible effects of differences in the social acceptability of items on the test scores by use of the forced-choice technique [8, 22, 25, 31, 32]. On the basis of available evidence Cronbach [4] has concluded that the forced-choice technique can be made relatively free from the influence of extraneous response sets, offering the possibility of increasing the predictive value of tests in which it is used. The technique involves the "grouping of alternative choices to make them appear of equal value and have unequal significance" [1, p. 424]. In a forced-choice scale of anxiety *S*s might be required to choose between self-descriptive statements presumed to be relevant to degree of anxiety and nonanxiety statements having equal value with respect to social favorability. The present investigation was concerned with the construction and evaluation of such a forced-choice scale of anxiety.

Procedure

Construction of the forced-choice anxiety scale. Two sets of anxiety items were used in the forced-choice scale. One set, hereafter referred to as FC-1, contained the 50 items constituting the *A* scale. The other set of 50 items, hereafter referred to as FC-2, contained MMPI items not originally selected as anxiety indicators that showed a correlation of .41 or more with the total *A*-scale score. The combined sets of 100 forced-choice items will be referred to as FC-T. Items were defined as nonanxiety statements if they were rated by none or only one of Taylor's five judges as descriptive of manifest anxiety and if they showed a correlation of .24 or less with the total *A*-scale score.[4]

---

[1] Modification of a dissertation submitted to the faculty of the Department of Psychology of the State University of Iowa in partial fulfillment of the requirements for the Ph.D. degree. The writer is greatly indebted to Drs. H. P. Bechtoldt and I. E. Farber, under whose direction this investigation was completed.

[2] Now at the Child Guidance Center, Des Moines, Iowa.

[3] Cook, W. C., Wood, E. C., McAllister, J., & Zide, N. The favorability ratings by anxious and nonanxious subjects of the items in the Iowa Biographical Inventory. Unpublished manuscript, State Univer. of Iowa, 1949.

In order to obtain favorability values for the anxiety statements and for the nonanxiety statements to be grouped with them, the 566 items in the group form of the MMPI and 34 items in the Wesley Rigidity Scale [30], a total of 600 statements, were rated on a five-point scale of social favorability by 108 students in introductory psychology. The *S*s were instructed to assume that the answer to each of the items was "yes." A rating of one represented a judgment of *very favorable* and five was *very unfavorable*. The mean favorability ratings of the 24 anxiety items that were stated negatively (e.g., "I don't usually worry") ranged from 1.38 to 2.44, while the mean favorability ratings of the 76 anxiety items that were stated positively (e.g., "I am inclined to worry") ranged from 2.72 to 4.62.[5] Thus the denial of the anxiety was consistently rated as more desirable than its admission.

Among the factors to be considered in constructing and administering a forced-choice scale is the number of choices in a block. Stewart [25] has reported that resistance is encountered when *S*s are required to choose from among uniformly unfavorable self-descriptive statements. She suggested that this resistance may be overcome if both desirable and undesirable alternatives are included in a forced-choice block consisting of three or more choices, and if *S*s are instructed to select a statement that is most descriptive of themselves and a statement that is least descriptive. Accordingly, each block in the present scale consisted of three statements: an anxiety statement, a nonanxiety statement matched with it for social favorability, and a second nonanxiety statement differing in social favorability from the two matched statements. None of the mean favorability ratings of the two matched statements differed by more than .10, while the second nonanxiety statement differed from the anxiety item by .80 to 1.25. If the matched statements were relatively unfavorable, the third statement was more favorable; if the matched statements were relatively favorable, the third statement was less favorable. Thus, within each block *S*s were to choose among statements differing with respect to social acceptability and with respect to their designation of anxiety by indicating the item most descriptive and least descriptive of themselves.

The following triad is a typical example of a forced-choice block used in this investigation:

Statement	Mean favorability rating
A. I have strong political opinions.	2.77
B. I sometimes tease animals.	3.82
C. I am a high-strung person.	3.88

Item C is an anxiety statement matched for favorability with item B, a nonanxiety statement. Item A is a nonanxiety statement rated as more favorable than the two matched statements. The position of the statements within each block was varied by means of a table of random numbers.

Statements were matched for favorability only if their frequency of selection by *S*s as self-descriptive was approximately equal.[6] Several statements bearing on the same behavior (e.g., "I am a good mixer" and "I love to go to dances") and mutually exclusive statements (e.g., "I almost never dream" and "I dream frequently") were not used in the same forced-choice block. In addition, items that appeared to be predominantly statements of fact rather than reflections of attitudes or judgments (e.g., "In school I was sometimes sent to the principal for cutting up") were not included in the forced-choice scale. Some statements (e.g., "I am a special agent of God") were excluded because it was believed that college students might not take them seriously.[7]

Scoring procedure. Two different scoring procedures were used. In one scoring method (Key 1) each of the three statements in a block was considered. Marking a positively worded anxiety statement as most self-descriptive was scored 1, and marking it as least self-descriptive was scored 0. When the two nonanxiety statements were marked and the anxiety statement was left unmarked, the anxiety statement was ranked between the two nonanxiety statements as being neither most nor least descriptive, so that the anxiety score was .5. When the anxiety statement was stated negatively, marking it as least self-descriptive was scored 1, and marking it as most self-descriptive was scored 0; leaving the anxiety statement unmarked was scored .5. By assigning scores of 0, .5, and 1, the possible range of scores was 50 for both FC-1 and FC-2, so that the Key 1 scores covered the same range as the scores on the original *A* scale.

---

[4] The relevant correlational data were obtained from an unpublished study by H. P. Bechtoldt, State Univer. of Iowa.

[5] The 600 items that were rated for social favorability, the instructions to the raters, and the mean favorability ratings are presented in the original dissertation on file in the State Univer. of Iowa library and in University Microfilms.

[6] The probability of selection has been indicated in several studies showing the probability of "yes" responses to MMPI items and to items in the Wesley Rigidity Scale. Personal communications from Dr. L. E. Drake, University of Wisconsin, and from Drs. W. G. Dahlstrom and H. P. Bechtoldt, State University of Iowa.

[7] The forced-choice scales and instructions are presented in the original dissertation on file in the State Univer. of Iowa library and in University Microfilms.

The MMPI group form numbers of the anxiety statements in FC-1: 7, 13, 14, 18, 23, 31, 32, 43, 67, 86, 107, 125, 142, 158, 163, 186, 190, 191, 217, 230, 238, 241, 242, 263, 264, 287, 301, 317, 321, 322, 335, 337, 340, 352, 361, 371, 397, 407, 418, 424, 431, 439, 442, 499, 506, 523, 528, 530, 549, 555.

The MMPI group form numbers of the anxiety statements in FC-2: 3, 8, 51, 54, 57, 79, 94, 97, 102, 122, 138, 146, 147, 148, 171, 172, 189, 201, 226, 234, 236, 266, 267, 279, 284, 299, 304, 305, 309, 336, 344, 356, 359, 374, 379, 382, 384, 389, 401, 402, 404, 411, 414, 416, 421, 425, 479, 509, 541, 560.

In the second scoring method (Key 2) all items scored 0 by the first method remained 0, and those scored 1 remained 1. When a positively worded anxiety item was left unmarked, a score of 1 was assigned if the matched nonanxiety statement was marked as least descriptive, and a 0 score was given if the matched nonanxiety statement was marked as most descriptive. When a negatively worded anxiety item was left unmarked, an answer of least descriptive for the matched nonanxiety statement was scored 0, and a marking of most descriptive was scored 1. The possible range of Key 2 scores was also 50 for both FC-1 and FC-2. Both scoring keys provided three anxiety scores, one for FC-1, another for FC-2, and a third for FC-T, the sum of FC-1 and FC-2.

Subjects. The forced-choice scale was administered to 209 *S*s in introductory psychology who had not made the favorability ratings. Each of these 209 *S*s had taken the Iowa Biographical Inventory, which included the *A* scale, about ten weeks prior to the administration of the forced-choice scale.

Results and Discussion

Relation between the forced-choice scale and the A *scale.* It was reasoned that if the correlations between the forced-choice form of the anxiety scale and the *A* scale did not differ significantly from the test-retest correlation obtained for the *A* scale, it might be concluded either that no extraneous source of variance due to the favorability factor is present in the *A* scale, or that the forced-choice form has not led to its elimination. Table 1 presents the correlations between the *A* scale and the Key 1 and Key 2 scores of FC-1, FC-2, and their total, FC-T.

Table 1

Correlations between the Taylor *A* Scale and Two Sets of Forced-Choice Items Treated Separately (FC-1 and FC-2) and in Combination (FC-T) on the Basis of Two Scoring Keys ($N = 209$)

Correlated Scales	Key 1	Key 2
A, FC-1	.73	.60
A, FC-2	.68	.58
A, FC-T	.74	.62

The correlation between the *A* scale and the Key 1 score of FC-1, which consisted of the *A*-scale items in forced-choice form, did not differ significantly from the previously reported [7] *A*-scale test-retest correlation of .83 ($t = 1.79$, $df = 268$). However, the correlation between the *A* scale and the Key 2 score of FC-1 was significantly lower ($p <$ .01) than the *A*-scale test-retest correlation ($t = 3.41$, $df = 268$). Furthermore, the correlation between the *A* scale and the Key 2 scoring of FC-1 was significantly lower ($R =$ .001, $t = 4.12$, $df = 206$) than the correlation between the *A* scale and the Key 1 scoring of FC-1.

In the light of other data to be presented it seems reasonable to interpret these results as indicating that the effects of social desirability may have been reduced or eliminated entirely as a source of variance common to the *A* scale and to the forced-choice scale only when Key 2 was applied to the responses in the forced-choice scale.

Relations between forced-choice subscales and scoring methods. It will be recalled that the anxiety items used in FC-2 were selected on the basis of their high correlation with the *A* scale when presented in single-choice form. A cross validation of these relationships could be obtained by correlating the FC-1 and FC-2 scores in the present study.

The correlation between FC-1 and FC-2, as scored by Key 1, was found to be .79. A correlation of practically equal magnitude (.74) was round when these two sets of items were scored by Key 2. In addition, the correlations between each of these two sets of items in forced-choice form and the *A* scale, as shown in Table 1, did not differ significantly, either for Key 1 scores ($t = 1.71$, $df = 206$) or for Key 2 scores ($t = .54$, $df = 206$).

These results may be taken as evidence that a set of items selected on the basis of high correlation with the *A* scale in the single-choice form were also correlated highly in the forced-choice form. Thus, scores on the forced-choice form appeared to be a function of factors similar to those influencing scores on the single-choice form of the test. Since both FC-1 and FC-2 could therefore be regarded as measures of manifest anxiety, these two sets of 50 forced-choice blocks were combined into a single scale of 100 blocks (FC-T).

Reliability. In order to evaluate the reliability of the forced-choice scales the internal consistency estimates of reliability were obtained by the use of the Kuder-Richardson formula 21 [13] and are reported in Table 2.[8]

Table 2

Reliability Coefficients of Three Sets of Forced-Choice Items Based on Two Scoring Keys ($N = 209$)

Set of Items	Key 1*	Key 2†
FC-1	.82	.69
FC-2	.80	.70
FC-T	.89	.83

* Kaitz modification.

† Kuder-Richardson formula No. 21.

The Kuder-Richardson estimate of reliability of the *A* scale, as computed for the sample of 209 *Ss* who also took the forced-choice scale, was .85. The estimated reliabilities of the Key 1 scores of FC-1 and FC-2 were not significantly lower than .85. The significant differences between the reliability coefficients involved the comparisons between the *A* scale and Key 2, and between the two scoring keys. The *A* scale was significantly more reliable ($p < .001$) than the Key 2 scores of FC-1 ($t = 5.14$) and of FC-2 ($t = 5.08$). The Key 1 forced-choice scale scores were more reliable ($p = .001$) than the Key 2 scores for FC-1 ($t = 7.31$), for FC-2 ($t = 8.56$), and for FC-T ($t = 4.25$). Thus the Key 2 scoring of the forced-choice scale yielded the least reliable measures.

The lower reliability of Key 2 as contrasted to the *A* scale might be accounted for in terms of a reduction in social favorability as a factor common to the responses to the forced-choice form. The lower reliability of the Key 2 scores as compared to the Key 1 scores might be due to two factors: a decreased influence of social favorability on the scores obtained by Key 2, and/or the reduction in the number of scoring categories from three in Key 1 to two in Key 2, a reduction which would probably decrease the discrimination among subjects.

Table 3

Mean, Median, Standard Deviation, and Range of the *A* Scale and of FC-1, FC-2, and FC-T, Based on Two Scoring Keys ($N = 209$)

Scale	Mean	Median	*SD*	Range
A scale	13.68	13	7.66	1–34
FC, Key 1				
FC–1	22.02	22.5	6.32	7.5–37.0
FC–2	23.06	23.5	6.14	8.0–39.5
FC–T	45.08	45.5	11.67	16.5–76.0
FC, Key 2				
FC–1	25.58	26	6.22	9–42
FC–2	26.91	27	6.30	9–42
FC–T	52.49	54	11.82	12–78

Distributions of scores. Table 3 shows the means, medians, standard deviations, and ranges of the scores for the *A* scale and for the forced-choice scale. The distribution of the *A*-scale scores was shown to be positively skewed by a test of normality given by McNemar [15], the deviation from normality being significant at the .02 level of confidence ($t = 2.35$). The distribution of FC-T scores was normal for Key 1, and negatively skewed at the .05 level of confidence ($t = 2.03$) for Key 2. Since the standard deviations of the *A* scale and of the Key 1 and Key 2 scores of the forced-choice scale were approximately the same, the elimination of the positive skewness characteristic of the *A* scale might be taken as an indication of the decreased influence of the social favorability factor on the forced-choice scale scores.

---

[8] In comparing the reliability of the *A* scale with that of the forced-choice scale the number of scoring categories for each item was considered. Both the *A* scale and the forced-choice scale, when scored by Key 2, allotted scores of zero or unity to the responses to a single item, so that the Kuder-Richardson estimate of reliability could be computed on the basis of 50 items. The scoring of the forced-choice scale by Key 1, however, involved three response categories which were assigned scoring weights of 0, .5, and 1. Accordingly, a modification of the Kuder-Richardson estimate of reliability developed by Kaitz [12] was utilized as more appropriate for Key 1 scores than the Kuder-Richardson formulas.

Table 4
Correlation between the MMPI *K* Scale and Two Sets of Forced-Choice Items Treated Separately (FC–1 and FC–2) and in Combination (FC–T) on the Basis of Two Scorings Keys ($N = 209$)

Correlated Scales	Key 1	Key 2
K, FC–1	–.55	–.36
K, FC–2	–.57	–.41
K, FC–T	–.57	–.42

Relations with the K *scale.* As the *K* scale of the MMPI was developed to serve as an estimate of the influence of social favorability on test scores, the magnitude of the relationship to the *K* scale of scores obtained on the *A* scale and on the forced-choice scale might also be used as a measure of sensitivity to the social favorability variable. In Table 4 is shown the relationship of the *K* scale with the *A* scale and with the forced-choice scale.

The negative correlations between the *K* scale and the Key 1 and Key 2 scores of FC-T were significantly lower ($p < .001$) than the correlation of $-.74$ between the *A* scale and the *K* scale ($t = 5.04$ for Key 1, 7.85 for Key 2, $df = 206$).[9] Thus the use of the forced-choice scale scored by either key resulted in a marked reduction in the influence of the social acceptability as measured by the correlation of the test scores with the *K* scale. This evidence of a reduction in the influence of social favorability was also found when FC-1 and FC-2 were considered separately.

The negative correlation between the Key 2 score of FC-T and the *K* scale was significantly lower ($p < .001$) than the correlation between the Key 1 score of FC-T and the *K* scale ($t = 5.90$, $df = 206$). Thus the Key 2 scores appeared to be significantly less susceptible to the influence of social desirability than the Key 1 scores.

The correlation between the Key 2 score of the forced-choice scale and the *K* scale still differed significantly from 0 ($p < .001$), which may be accounted for, in part, by the overlap in the forced-choice scale and in the *K* scale of 11 statements that contribute to the anxiety score when answered in one way and to the *K*-scale score when answered in the other.

Bias experiment. The most direct method of evaluating the influence of a social desirability factor upon the various anxiety measures is to instruct *S*s to answer deliberately in such a way as to appear in the most favorable light possible. A number of studies [2, 11, 21] have shown that scores on personality tests of the "yes-no" type are greatly influenced by such instructions. On the basis of the reasoning underlying the construction of the forced-choice scale, it would be predicted that it would not be so greatly influenced by biasing instructions.

Sixty-four *S*s from introductory psychology, who had not previously made favorability ratings or taken the forced-choice scale, were divided into two groups matched on the basis of their *A*-scale scores. About two weeks following the administration of the *A* scale one group of 32 *S*s took the *A* scale under instructions to appear in the most favorable light possible, and the other group of 32 *S*s took the forced-choice scale under similar instructions. In order to evaluate the influence of biasing instructions it was necessary to express the raw scores in comparable units. Accordingly the forced-choice scale scores and *A*-scale scores were transformed linearly into a distribution having a mean of 20 and a standard deviation of 4. The coded means, medians, and standard deviations obtained by the matched groups of 32 *S*s under conventional and biasing instructions are shown in Table 5. The following discussion is based on the coded scores.

The mean of the faked *A*-scale scores was significantly lower, at the .001 level, than the mean of the conventional *A*-scale scores ($t = 6.31$, $df = 31$). The mean biased score of FC-T, Key 1, was also significantly lower, at the .01 level, than the mean conventional *A*-scale score ($t = 2.80$, $df = 31$). On the other hand, the mean biased score of FC-T, Key 2, did not differ significantly either from the mean of the conventional *A*-scale scores ($t = 1.78$, $df = 31$) or from the mean conventional Key 2 forced-choice score of 20.00 of the standard sample of 209 *S*s. These results demonstrate that the *A* scale and the Key 1 scoring of the forced-choice scale are subject to the influence of the social favorability variable, while the Key 2 scores appear to be relatively insensitive to defensive attitudes.

Furthermore, the mean biased *A*-scale score was significantly lower ($p < .01$) than the mean biased score of FC-T, Key 1 ($t = 3.60$, $df = 31$). In turn, the mean bias score of FC-T, Key 1, was significantly lower ($p < .001$) than the mean bias score of FC-T, Key 2 ($t = 5.53$, $df = 31$). It is

---

[9] The correlation between the *A* scale and the *K* scale was based on the responses of *S*s who took the *A* scale about ten weeks prior to the administration of the *K* scale. The *K* scale was part of a battery of tests that was administered jointly with the forced-choice scale. It is not believed that the time interval affected the correlation between the *A* scale and the *K* scale to any marked degree.

Table 5
Transformed Means, Medians, and Standard Deviations of Conventional and Biased *A*-Scale Scores, and of Biased Forced-Choice Scale Scores (FC-T) Scored by Two Keys for Two Groups of 32 Matched *Ss*

	A-scale group		Forced-choice scale group		
	A-scale conv.	*A*-scale bias	*A*-scale conv.	FC-T, Key 1 bias	FC-T, Key 2 bias
Mean	19.13	15.03	19.03	17.18	20.26
Median	18.08	13.90	17.82	17.53	20.02
SD	3.29	2.49	3.74	1.92	2.57

apparent that the forced-choice scale, scored by either key, was less sensitive to deliberate faking than the *A* scale. Nevertheless, forced-choice scale scores were more susceptible to faking when the second nonanxiety statement was considered in the Key 1 scoring.

Additional considerations. The instructions used in the bias experiment were equivalent to instructions to rate items for favorability. Scoring by Key 2, under biasing instructions, provided a means of estimating whether statements of equal social favorability when rated singly were rated as equally desirable after they were grouped into forced-choice blocks. If the matched anxiety and nonanxiety statements were regarded as equally favorable, one might expect that each of the two statements would have an equal chance of being selected as the more favorable, so that on the average the mean of the faked Key 2 scores of the forced-choice scale of 100 blocks would be 50. The mean faked Key 2 score of FC-T, not coded, was 53.22, which differed from a hypothetical mean of 50 between the .02 and .05 levels of confidence ($t = 2.36$, $df = 31$). Thus the anxiety statements were selected somewhat more frequently as being desirable than their matched nonanxiety statements. This circumstance may account for the slight negative skewness of the distribution of the Key 2 scores. It may be concluded that the procedure of matching statements on the basis of their individual favorability ratings did not insure exactly equal favorability when the matched statements were put into forced-choice form.

The effectiveness of a forced-choice scale of anxiety in reducing or eliminating social favorability as a variable affecting test scores may also be limited if anxious and nonanxious individuals differ in their estimate of the social favorability of anxiety statements. In order to test this possibility 24 anxious and 19 nonanxious *Ss* were selected from the 108 *Ss* who had rated the MMPI and the Wesley Rigidity Scale items for social favorability. Individuals who scored in the upper 20% of the *A* scale were defined as anxious, and those who scored in the lower 20% were defined as nonanxious. The favorability ratings of these two groups of *Ss* were compared with respect to the 50 anxiety items from the *A* scale and their matched nonanxiety statements. The 37 positively stated anxiety items in FC-1 were considered separately from the 13 negatively stated items.[10]

The difference scores between the favorability ratings of the anxious and nonanxious *Ss* were computed for each of the 37 positively stated anxiety items and for each of their matched nonanxiety items. The anxious *Ss* rated these 37 anxiety statements as significantly more favorable, at the .001 level, than did the nonanxious *Ss* ($t = 4.70$, $df = 72$). Applying the same procedure to the 13 negatively stated anxiety items and to their matched nonanxiety items, it was found that anxious *Ss* rated the anxiety items as significantly less favorable ($p < .001$) than did the nonanxious *Ss* ($t = 5.97$, $df = 24$). These results indicate that anxious *Ss* regard the expression of anxiety less unfavorably than do nonanxious *Ss*.

Since anxious *Ss* differed from nonanxious *Ss* in their favorability ratings of anxiety statements, the influence of social favorability was probably not entirely eliminated from the test scores regardless of the form of presentation and of the scoring method. However, the results of this investigation indicate that the influence of social desirability on test scores could at least be drastically reduced by the use of the forced-choice method.

---

[10] Favorable ratings of a negatively worded item had the same meaning as unfavorable ratings of a positively worded item. For instance, when *Ss* gave a relatively favorable rating to a negatively worded item, such as "I don't usually worry," they considered the behavior of worrying as undesirable.

Summary

The present study was concerned with the development of a forced-choice form of the Taylor Anxiety Scale designed to reduce the effects of possible tendencies by *S*s to consider the social desirability of particular responses. The forced-choice form consisted of three statements, an anxiety statement and a nonanxiety statement of comparable social favorability, and a second nonanxiety statement differing in social favorability from the two matched statements. Favorability ratings were obtained for 566 MMPI items and for 34 items from the Wesley Rigidity Scale. In 50 forced-choice blocks (FC-1) the items constituting the original *A* scale were used as anxiety statements, and in 50 forced-choice blocks (FC-2) items which exhibited a high correlation with total scores on the original *A* scale were used as anxiety statements. The *S*s were required to choose one statement in each triad as most descriptive and one as least descriptive of themselves. By use of Key 1, scores were assigned according to the ranking of the anxiety statement as most or least descriptive relative to the two nonanxiety statements, while in Key 2 the ranking of the anxiety statement relative only to its matched nonanxiety statement was scored. This forced-choice scale of anxiety was administered to 209 *S*s who had taken the *A* scale about ten weeks previously.

The following results were obtained:

1. The correlation between FC-1 and the *A* scale was significantly lower that the *A*-scale test-retest correlation when Key 2 was used, but not when Key 1 was used.

2. Items selected on the basis of their high correlation with the *A* scale in the single-choice form were also correlated highly in the forced-choice form. The scores obtained on FC-1 and FC-2 could therefore be combined into FC-T.

3. The *A* scale and the Key 1 score of the forced-choice scale were more reliable in terms of internal consistency than the Key 2 score of the forced-choice scale.

4. The positive skewness of the distribution of the *A*- scale scores was corrected when the forced-choice scale was used.

5. The negative correlation with the *K* scale of the MMPI was highest for the *A* scale and lowest for the Key 2 score of the forced-choice scale. Both differed significantly from the Key 1 score of the forced-choice scale.

6. One group of 32 *S*s took the *A* scale under instructions to appear as favorable as possible, and another group of 32 *S*s, matched with the first group on the basis of their *A*-scale scores, took the forced-choice scale under the same biasing instructions. While both the *A* scale and the Key 1 scores of the forced-choice scale were lowered by deliberate faking, this reduction was greater for the *A* scale. The Key 2 scores of the forced-choice scale were not affected by biasing instructions.

7. Statements matched on the basis of social favorability were not of equal value after they were put into forced-choice form.

Anxious subjects differed from nonanxious subjects in their estimate of the social desirability of anxiety statements. This was taken as evidence that the influence of social favorability may not have been entirely eliminated by the use of the forced-choice technique. However, the results of this investigation indicated that the influence of social desirability on test scores could at least be drastically reduced.

Received March 23, 1953.

References

1. Baier, D. E. Reply to Travers' "A critical review of the validity and rationale of the forced-choice technique." *Psychol. Bull.*, 1951, **48**, 421-434.
2. Bernreuter, R. G. Validity of the personality inventory. *Personnel J.*, 1933, **11**, 383-386.
3. Cameron, N. *The psychology of behavior disorders.* Boston: Houghton Mifflin, 1947.
4. Cronbach, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1950, **10**, 3-31.
5. Farber, I. E., & Spence, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120-125.
6. Gleser, Goldine, & Ulett, G. The Saslow screening test as a measure of anxiety proneness. *J. clin. Psychol.*, 1952, **8**, 279-283.
7. Hedlund, J. L., Farber, I. E., & Bechtoldt, H. P. Normative characteristics of the manifest anxiety scale. *Amer. Psychologist*, 1951, **6**, 504. (Abstract)
8. Highland, R. W., & Berkshire, J. R. A methodological study of forced choice performance rating. Research Bulletin 51-9, Human

Resources Research Center, Lackland AFB, May, 1951.

9. Hilgard, E. R., Jones, L. V., & Kaplan, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94-99.

10. Holtzman, W. H., Calvin, A. D., & Bitterman, M. E. New evidence for the validity of Taylor's Manifest Anxiety Scale. *J. abnorm. soc. Psychol.*, 1952, **47**, 853-854.

11. Hunt, H. F. The effect of deliberate deception on Minnesota Multiphasic Personality Inventory performance. *J. consult. Psychol.*, 1948, **12**, 396-402.

12. Kaitz, H. B. A note on reliability. *Psychometrika,* 1945, **10**, 127-131.

13. Kuder, G. F., & Richardson, M. W. The theory of the estimation of test reliability. *Psychometrika,* 1937, **2**, 151-160.

14. Lucas, J. D. The interactive effects of anxiety, failure, and intraserial duplication. *Amer. J. Psychol.*, 1952, **56**, 54-66.

15. McNemar, Q. *Psychological statistics.* New York: Wiley, 1949.

16. Meehl, P. E., & Hathaway, S. R. The *K* factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *J. appl. Psychol.*, 1946, **30**, 525-564.

17. Montague, E. K. Role of anxiety in serial rote learning *J. exp. Psychol.*, 1953, **45**, 91-96.

18. Peck, Ruth. Influence of anxiety factors upon the effectiveness of an experimental "counseling" session. Unpublished doctor's dissertation, State Univer. of Iowa, 1951.

19. Ramond, C. Anxiety and task as determiners of verbal performance. *J. exp. Psychol.*, 1953, **46**, 120-124.

20. Rosenbaum, G. Stimulus generalization as a function of level of experimentally induced anxiety. *J. exp. Psychol.*, 1953, **45**, 35-43.

21. Ruch, F. L. A technique for detecting attempts to fake performance on a self-inventory type of personality test. In Q. McNemar & M. A. Merrill (Eds.), *Studies in personality.* New York: McGraw-Hill, 1942. Pp. 229-234.

22. Sisson, E. D. Forced-choice—the new Army rating. *Personnel Psychol.*, 1948, **1**, 365-381.

23. Spence, K. W., & Farber, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116-119.

24. Spence, K. W., & Taylor, Janet A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183-188.

25. Stewart, Naomi. Methodological investigation of the forced-choice technique, utilizing the officer description and the officer evaluation blanks. AGO, Personnel Research Section, Report No. 701, 6 July 1945.

26. Taylor, Janet A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81-92.

27. Taylor, Janet A. A personality test for manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285-290.

28. Taylor, Janet A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61-66.

29. Wenar, C. Reaction time as a function of manifest anxiety and stimulus intensity. Unpublished doctor's dissertation, State Univer. of Iowa, 1951.

30. Wesley, Elizabeth L. Perseverative behavior in a concept-formation task as a function of manifest anxiety and rigidity. *J. abnorm. soc. Psychol.*, 1953, **48**, 129-134.

31. Wherry, R. J. Comparative validity of the WDAGO form No. 67 and the FCL-2 according to various breakdowns. AGO Personnel Research Section, Report No. 671, 7 Dec. 1945.

32. Wherry, R. J. Comparative validity of the WDAGO form No. 67 and the FCL-2 according to various breakdowns II, European Theater. AGO Personnel Research Section, Report No. 673, 8 Dec. 1945.

Journal of Consulting Psychology
Vol. 17, No. 6, 1953

A Transposed Factor Analysis of Paranoid Schizophrenics[1]

Wilson H. Guertin
Veterans Administration Hospital, Knoxville, Iowa

and Victor Zilaitis
Beatty Memorial Hospital, Westville, Indiana

In an earlier study, Guertin [3, 4] found some factor analytic evidence for three types of paranoid schizophrenics, but the general domain of schizophrenia investigated was too broad to bring the smaller area of paranoid schizophrenia into clear focus. The three tentative factors suggested were the *persecuted-suspicious,* the *superior-grandiose,* and the *over-ideational.* In a transposed factor analysis of a portion of these data [5] a large first factor was obtained which could be identified as paranoid schizophrenia but again the sample was too broad to provide a focus solely on paranoid schizophrenia.

The present report is on a propaedeutic study designed to explore the possibility of further subtyping paranoid schizophrenics by sampling the behavior of only paranoid schizophrenics and subjecting the results to a transposed factor analysis.

Method

The study is based upon 24 hospitalized males who were diagnosed as paranoid schizophrenics by the hospital staff. In all cases the investigator looked for intrinsic or extrinsic evidence of paranoid psychosis at the time of testing in order to avoid including any subjects who might be in complete remission. The twenty-fifth subject is a hypothetical "normal" individual derived from normative data in the MMPI manual.

It is interesting that an earlier study independently employed this "normal" individual in a similar fashion [1]. It was necessary to exclude over one-half of the original paranoid sample selected for this study for various reasons. Among the original 51 subjects, only two had to be discarded because they were uncooperative and hostile. This would suggest that there was an adequate sampling of hostile-reacting individuals. Five were excluded because they appeared to be in good remission while four had to be discarded because of foreign-language background which made communication difficult. Two more individuals were discarded because their distribution of true and false answers left too small a proportion in one of the tails. It was felt that no serious harm was done to the representativeness of the sample because of these exclusions. However, it was necessary to exclude 13 individuals who were too confused as demonstrated by their clinical behavior or their unreliability on the test. This suggests that the only really important biasing of the final sample was in the direction of the exclusion of paranoids with appreciable confusion.

It was necessary to employ some technique of evaluating the individuals on a number of test items in order to make a transposed analysis. The choice lay among three areas: enumeration of symptoms, checklist of behavior, and avowed attitudes and interests. Since the MMPI was designed to include sampling of significantly deviant paranoid behavior, it was felt that these cards might provide a group of items which would prove useful for sampling the avowed attitudes and interests of the paranoid schizophrenics. The authors independently, and later jointly, went through the total 550 MMPI cards and selected those items which they agreed probably would be most discriminating among different types of paranoid schizophrenics. In general, there was good agreement on these selections and both investigators concurred independently on the final 100 items employed. An additional fifteen items from the *L* scale of the MMPI were included to provide some statistical control over the distortion of answers.

The subjects were told that this was a research study and that their assistance would be appreciated. They were assured that the information obtained was not for their hospital folders but would be used for scientific purposes. Each subject was asked to sort the 115 cards into the true and false boxes according to whether he felt that the statement was true for him or not. All subjects, with the exception of two, complied very well with this and it was unnecessary to do much forcing of items into one box or another. After the 115 items had been sorted, each subject was asked to re-sort the 15 items from the *L* scale in order to provide a simple reliability check. With these controls it was possible to say that the results were reliable and that some external criterion was available to indicate the validity of the sorting of the 100 test items. The mean number of *L* items answered in the proper direction was 10.48, thereby indicating relatively little distortion operating for this sample. The mean number of agreements upon retest with the *L* items of 13.24 is also high.

---

[1] This work was conducted at Beatty Memorial Hospital, and the report completed at and released from Veterans Administration Hospital, Knoxville, Iowa.

Tetrachoric intercorrelations among individuals were obtained from fourfold contingency tables and charts. The resulting intercorrelation table was factored by the multiple-group centroid method. Communalities were estimated by employing the largest correlation within a cluster for those individuals within a cluster. Those who lay outside of clusters had communality estimates based upon the highest intercorrelation in the whole column.

Results

The intercorrelation matrix, which will not be presented here in order to conserve space, disclosed a great many medium high as well as high correlations with positive signs. Negative signs were largely absent. In doing a cluster analysis, it became obvious that one very large group factor was predominant. After extracting two group factors simultaneously, it was decided that a complete centroid check should be made to see whether the first of these oblique factors had pulled out sufficient variance. A complete centroid through the reduced intercorrelation matrix extracted a factor which accounted for 55% of the estimated communality. In contrast to this, the first group centroid factor had accounted for 45% of the estimated communality. It was felt that this group factor would be more meaningful and accounted for sufficient communality to be retained. Further cluster analysis was made on the residual matrix obtained after the first two oblique factors were extracted. After four identifiable clusters had been located, the final factoring was begun. Mutiple-group centroid extraction of these four factors simultaneously produced the oblique factor matrix reported in Table 1. Inasmuch as no further clusters of individuals could be determined and since 85% of the estimated communality had been already extracted, it was decided to discontinue the factoring at this point. The clustering of the individuals in a relatively small "test space," which was determined predominantly by the first oblique factor, is seen in the intercorrelations among the factors. Only the correlation between factors *B* and *D* was negligible. The correlations of factors *C* and *D* with *A* were of the order of .6, while correlations of the order of .4 were found between factors *A* and *B*, *B* and *C*, and *C* and *D*. Subject No. 25,

Table 1

Factor Matrices for Subjects

Subj. no.	Oblique factors*				Orthogonal representation			
	A	B	C	D	I	II	III	IV
17	*88*	36	58	52	88	00	05	00
25	*84*	40	78	46	84	06	33	−06
12	*83*	70	75	37	83	39	21	−13
9	*80*	22	49	45	80	−12	03	−04
21	*73*	32	56	32	73	02	14	−15
6	*68*	39	27	53	68	12	−22	19
15	*53*	38	49	45	53	18	17	18
4	*53*	21	32	44	53	−01	00	16
3	*51*	32	15	45	51	12	−24	17
2	30	*80*	39	41	30	74	08	36
13	26	*75*	47	−04	26	71	22	−18
23	36	*71*	11	01	36	62	−30	−17
14	29	*58*	53	37	29	51	33	28
16	23	*47*	27	22	23	41	06	25
1	05	*34*	30	06	05	35	−09	05
8	64	56	*83*	24	64	33	45	−19
24	27	18	*79*	11	27	08	80	−11
5	72	56	*72*	48	72	29	29	08
11	46	62	*69*	51	46	47	41	32
22	57	12	*65*	22	57	−13	43	−19
20	33	31	*49*	23	33	19	33	04
7	43	12	35	*71*	43	−06	13	55
18	26	06	22	*67*	26	−05	09	63
10	47	15	21	*59*	47	−05	−09	29
19	12	28	25	*38*	12	25	17	40
				$\Sigma a_{ij}^2 =$	7.19	2.76	2.02	1.64

* Decimal points omitted.

who was the normative one obtained from the MMPI, shows interesting loadings on the factors and determines to a large extent the identity of Factor *A*.

Discussion

The estimated communality among the paranoid schizophrenics in this sample of 63.3% of the total variability of all the individuals' test behavior gives good support for the commonness of people within this diagnostic category. It is seen from this statistic that they are more similar in behavior on these test items than they are different. The four group factors extracted accounted for 85% of this communality and 54.4% of the total variance of the individuals. This means that by reference to these four personality types, more than 50% of the variability of these individuals can be described.

Socially Normal Paranoid Type (A)

This is the type of paranoid who makes a good social impression in his behavior and makes a diagnosis of paranoid schizophrenia difficult. Usually, diagnoses of such a paranoid psychosis depend upon obtaining a valid history of delusional action in a single rather restricted area. This factor or personality type might be described as a normalcy factor inasmuch as individuals who are of this type correspond closely to the answers given by the normative individuals of the MMPI (subject No. 25). This factor seems to be largely a common species factor as described by Cattell [2], i.e., they are similar in that they are human beings. The only consistent deviant response they give is wondering if they locked the door and closed the windows when they are away from home. Among their normal responses, there is nothing particularly striking. Individuals who approximated this type, almost without exception, gave indications of restricted delusions on just a few of the 100 MMPI items. These few markedly abnormal responses appearing for a subject did not have sufficient weight to prevent him from being predominantly of this type.

The nine most typical type *A* individuals gave the deviant answer: worries about doors and windows. Nondeviant answers were: not happy alone, not an important person, not being talked about, never feels like smashing things, not obsessed with bad words, home life pleasant, no strange or peculiar sensations, no one stealing ideas, not afraid of knives, not being followed, most people would not lie to get ahead, not inspired to life of duty, never kept diary, and never feels like going to pieces

Grandiose and Delusional Paranoid Type (B)

This type of paranoid is a very outgoing and overproductive person whose delusional systems are largely in the grandiose area. People, to him, are objects of his environment without any emotional importance; overideational trends are also present for this type. These individuals gave a mean number of true responses for the 100 items of 65.5 which, as compared with the mean of the other three groups of only 41 responses, seems to be clear evidence of overproductiveness and expansiveness.

The four most typical type *B* individuals gave the deviant answers as follows: strong political opinions, agent of God, peculiar odors, being talked about, strange and peculiar sensations, unusual religious experiences, inspired to life of duty, never felt better, most people would lie to get ahead, likes to know who to get next to on a job, people make friends because they are useful, and wonders why people do things for him. Nondeviant answers were: daily life full of interesting things, talks to strangers, likes serious lectures, good leader, sex life O.K., likes science, and judgment never better.

Evasive and Well-Integrated Paranoid Type (C)

This type of individual seems to be one who retains a good integration and is able to conceal many of his delusions. Just what the type of delusion might be probably would vary among individuals but they all seem to be rather successful in covering up their areas of difficulty. One is impressed by the nature of their uniform denials on test items. Related to this is the fact that they gave fewer true answers than any of the other types. One gets the impression that this is the type of individual who is rather amoral and not too cooperative or productive in mental status inquiries. This type of individual would also make diagnosis difficult although the history should be more productive than for the type *A* who seems to have more restricted delusions. The relationship between this type and type *A* is extremely high, probably because they both retain a high degree of personality integration.

The five most typical type *C* individuals gave the deviant answer: would not like expensive clothes. Nondeviant answers were: doesn't want to know who to get next to on a job, don't feel like failure when hearing of another's success, never made to do things by hypnosis, not bothered by sex thoughts, safe to trust people, can be friendly with those who do wrong, never disgusted over sex, parents made him obey, never about to go to pieces, likes science, seldom constipated, and nobody influencing mind.

Sensitive, Inadequate, and Withdrawn Type (D)

This type of individual seems to experience extreme feelings of inadequacy and many neurotic-type conflicts. He seems to be an immature person who has tried unsuccessfully to please the parents and has now come to evaluate himself as a "bad boy." Inhibitions and suppression of overt expression are paramount in this type of individual. We see this in the area of sexual as well as other interpersonal behavior. This individual does not seem to be very delusional and it is seen that the correlation between the delusional type *B* and this one is quite low. It is questionable whether such individuals should be classified as paranoid inasmuch as they appear to manifest predominantly catatonic symptoms.

The three most typical type *D* individuals gave deviant answers: others know thoughts, strangers look critically, dirt disgusts, compulsive counting, obsessed with bad words, been in trouble over sex, bothered by sex thoughts, criticism hurts terribly, dislikes dances, never talks to strangers, dislikes serious lectures, dislikes science, people make friends because they are useful, people don't like to help others, and wonders why people help him. Non-deviant answers were: conduct controlled by customs, doesn't want to know who to get next to on a job, keeps mouth shut when in trouble, thoughts come too fast to speak, could never benefit world, have felt better than now, sleeps without disturbing thoughts, not afraid of germs, never had unusual religious experience, and never kept diary. There were too few individuals in the last classification to make a good evaluation of the nature of the factor type and yet one feels that they do not share the common features that the first three factor types do.

The present study is regarded as being too restricted to yield results which can be proposed as classificatory types or criteria, but should serve as an illustration of the usefulness of the transposed factor-analytic technique in deriving such systems. It was felt that the selected MMPI items proved to be very useful samplings of personality variables even though they have a subjective basis. It is the investigators' opinion that these items can be employed successfully in other factor-analytic studies of pathological groups that might be capable of responding reliably and validly, such as cases of organic brain disease. It does seem clear from these results that at least two subtypes of paranoid psychoses should be distinguished: the first, the socially normal well-integrated individual, and secondly, the grandiose delusional person. It is conceivable that these two phenotypes might represent different phases of the same disease process.

Summary and Conclusions

1. Much criticism has been made of Kraepelinian classification, leading to many constructive empirical investigations. The authors have felt that such empirical studies should be encouraged and that methodological experimentation will prove profitable in eventually establishing more effective classificatory systems.

2. This paper is a research report on a transposed factor-analytic study of 24 male paranoid schizophrenics and one statistically "normal." One hundred selected Minnesota Multiphasic items were employed to provide samples of avowed attitudes and interests. Brief checks upon reliability and validity were incorporated.

3. The factor analysis disclosed a large commonness among the individuals in this study. Three personality types included under the classification of paranoid schizophrenia are suggested. They are as follows: *A — socially normal paranoid, B — grandiose and delusional paranoid,* and C — *evasive and well-integrated paranoid.* A fourth factor type disclosed is the sensitive, inadequate, and withdrawn type. This latter may correspond to the catatonic subtype of schizophrenia.

4. In general there was a confirmation of the classification of paranoid, but at least two types are different enough to merit separate classificatory labels. They were the socially normal well-integrated person and the grandiose delusional type.

Received February 9, 1953.

References

1. Borko, H. A factor-analytic study of the Minnesota Multiphasic Personality Inventory using a transpose matrix (Q-technique). *Amer. Psychologist,* 1952, **7**, 342. (Abstract)
2. Cattell, R. B. *Factor analysis.* New York: Harper, 1952.
3. Guertin, W. H. A factor analytic study of schizophrenic symptoms. Unpublished doctor's thesis, Michigan State Coll., 1951.
4. Guertin, W. H. A factor analytic study of schizophrenic symptoms. *J. consult. Psychol.,* 1952, **16**, 308-312.
5. Guertin, W. H. An inverted factor analysis of schizophrenics. *J. consult. Psychol.,* 1952, **16**, 371-375.

Journal of Consulting Psychology
Vol. 17, No. 6, 1953

Validity of a Paper-and-Pencil Form of the MMPI Psychopathic Deviate Scale

Solomon Machover
Kings County Hospital and N. Y. State University College of Medicine

and Helen J. Anderson
Brooklyn College

It is occasionally of clinical interest to know how a patient performs on a single constituent scale of the Minnesota Multiphasic Personality Inventory. Under such circumstances the administration and scoring of the entire inventory seems an unwarrantedly long and cumbersome procedure. Furthermore, the conventional way of recording and scoring responses tends to deflect attention away from the content of the individual responses and from the patterns of responses which contribute to the total scale scores. Material of manifest and potential clinical significance is thus too readily overlooked. Both limitations on the usefulness of the MMPI would be obviated if there were available paper-and-pencil forms of the separate constituent scales of the inventory.

Preparation of forms for the separate scales involves the segregation of items pertaining to each of the scales. The subject is then required to respond to items which are in a sequence of related items and divorced from the total context of randomly distributed unrelated and "silent" items. What effect does this segregation and massing of items into separate constituent scales singly administered have on the validity of the resultant scores? The present study was designed to answer this question as it applies to the Psychopathic Deviate scale which has been reported [1] to have reasonably satisfactory validity when administered in the standardized card form.

Method

A mimeographed form of the Psychopathic Deviate scale was prepared with the 50 items of the scale arranged in the same order as in McKinley and Hathaway's original listing [3]. By way of identification, each item was lettered and numbered to accord with its designation in the manual [2]. To facilitate scoring, each item was assigned either an even or an odd number which was imprinted in front of the identifying designation so that a subject's score on the scale might be found by simply counting the even-numbered items marked "true" and the odd-numbered items marked "false." The mean and standard deviation for normal males were imprinted at the bottom of the form for use in conversion of raw scores into T scores. Instructions to the subject and the first two and last three items of the scale follow as samples to illustrate the arrangement.

Below are 50 statements about yourself. Read each statement carefully. If what it says is *true for you,* draw a circle around the T at the end of the line. If what it says is *false for you,* draw a circle around the F. Start at the top of the page. Mark all statements. Work fast.

1A4 I am neither gaining nor losing weight. T F
6B42 I have used alcohol excessively. T F

3I26 I am easily downed in an argument. T F
4I27 I find it hard to keep my mind on a task or job. T F
8I30 My way of doing things is apt to be misunderstood by others. T F

Mean 14.00 Standard Deviation 4.00

Subjects were 50 randomly selected patients admitted for observation to the psychiatric division of the Kings County Hospital. The median age of the subjects was 34 years, with a range of 16 to 59. They were classified psychiatrically as follows:

Chronic alcoholism	21
Drug addiction	2
Mental deficiency	1
Schizophrenia	9
Psychosis due to alcohol	4
Manic-depressive psychosis	4
Psychoneurosis	2
Psychopathic personality	7

Half the subjects were individually tested first with the paper-and-pencil massed form of the *Pd* scale and then, after one to three days, with the entire inventory in standard card form. The remaining half received the standard inventory first, followed, after a like interval, by the experimental form of the *Pd* scale. It was thus possible to check the validity of the experimental paper-and-pencil form of the *Pd* scale against the criterion of the *Pd* scale derived in standard fashion from the administration of the MMPI as a whole.

Results

As shown in Table 1, the mean *T* score of the 50 subjects on the experimental form of the *Pd* scale is 64.24 as against a mean of 63.56 on the standard form of the scale. The difference is well within chance expectancy, *p* being equal to .62 for a two-tailed test. Rotation of the order of administration to equalize for both forms any effect of practice or familiarity seems to have been superfluous. The difference between the means for tests administered first and for those administered second is exactly zero. Nor does it seem to matter whether the experimental form or the standard form is administered first.

The Pearson correlation coefficient between experimental-form scores and standard-form scores was found to be .79. In view of the fact that a test-retest reliability coefficient of .71 is reported by McKinley and Hathaway [3] for the standard form of the *Pd* scale, it may be said that the paper-and-pencil form of the *Pd* scale is as valid, against the criterion of the standard *Pd* scale, as the standard *Pd* scale is itself reliable. Greater validity, of this derived type, cannot be expected.

Discussion

It appears that no loss of accuracy results from segregating the items of the *Pd* scale and administering them in massed sequence as a separate paper-and-pencil scale out of the context of the MMPI as a whole. Such cushioning effect as may normally be exerted by the "silent" items of the MMPI and by items not relevant to the *Pd* scale seems to be without consequence for the total score on the scale. It seems impressive that the equivalence of the out-of-context *Pd* scale with the standard form

Table 1

Comparison of *T* Score Means on the Standard and Experimental Forms of the MMPI *Pd* Scale

Item	*N*	Range of *T* scores	Mean	*SD Diff.*	*t*	*p*
Standard *Pd* scale	50	37—98	63.56	1.33	.51	.62
Experimental *Pd* scale		35—93	64.24			
First-test scores	50	37—93	64.0	1.33	0	1.00
Second-test scores		35—98	64.0			
Experimental scale	25	42—93	67.24	2.18	.35	.71
Standard scale		45—98	66.48			
Standard scale	25	37—86	60.64	1.57	.38	.73
Experimental scale		35—88	61.24			

of the scale has been established with a quite deviate population. While direct evidence is lacking, this finding at least suggests the possibility that others of the constituent scales of the MMPI may also fare well as separate scales.

A number of advantages accrue to the availability of the constituent scales which may be administered in separate paper-and-pencil form. The clinician's attention is perforce directed to where it should be, viz., the subject's actual responses. Identical total scores on a given scale may stand for quite different patterns of actual behavior. Variations in item clusters, although irrelevant to total score, may be quite significant for personality evaluation and behavior prediction. Ready availability of the concrete responses provides provocative material which is useful in structuring follow-up clinical interviews. Finally, availability of the separate scales is a contribution to efficiency in those instances in which information on less than the total inventory is sought.

Summary

1. A paper-and-pencil form of the MMPI *Pd* scale, prepared as a separate scale, was administered individually to a group of 50 psychiatric patients. Scores on this experimental form of the *Pd* scale were compared with scores obtained by the same subjects on the *Pd* scale when administered in standard form as part of the total MMPI.

2. A correlation of .79 was obtained between scores on the experimental form and scores on the standard form of the *Pd* scale. The out-of-context paper-and-pencil *Pd* scale may be said to be as valid, against the criterion of the standard form of the scale, as the standard form of the scale is itself reliable.

3. The mean score for the experimental form was found to be not significantly different from the mean score for the standard form. Order of administration of the two forms made no difference.

4. Availability of a separate paper-and-pencil form permits more efficient and flexible use of the MMPI, directs attention to clinically significant content, presents data useful in follow-up interview, and provides a permanent record of detailed concrete responses in readily utilized form.

Received March 2, 1953.

References

1. Benton, A. L. The Minnesota Multiphasic Personality Inventory in clinical practice. *J. nerv. ment. Dis.*, 1945, **102**, 416-420.
2. Hathaway, S. R., & McKinley, J. C. *Manual for the Minnesota Multiphasic Personality Inventory*. New York: Psychological Corp., 1943.
3. McKinley, J. C., & Hathaway, S. R. The Minnesota Multiphasic Personality Inventory: V. Hysteria, hypomania and psychopathic deviate. *J. appl. Psychol.* 1944, **28**, 153-174.

Journal of Consulting Psychology
Vol. 17, No. 6, 1953

Ellis on Personality Inventories

Allen Calvin[1] and James McConnell
The University of Texas

In a recent article in this journal, Albert Ellis seriously questions the value of the personality inventory [2]. He raises a series of objections to the use of these inventories which do not seem fully justified to us. Although Ellis questions the value of all such inventories, he mentions only the MMPI by name. Since we cannot be certain which techniques other than the MMPI Ellis has reference to in his study, and since it is known that the various inventories show serious differences among themselves in construction, reliability, and validity, we will confine our rebuttal to a defense of the MMPI.

It is perhaps best that Ellis' sweeping condemnation of the personality inventories be answered point by point. His objections are centered around five arguments, the first of which is that "in most instances the inventories are not measuring the independent traits they are supposed to be measuring" [2, p. 48]. In support of this first point he cites 21 studies in which intercorrelations were calculated among the subscales of different inventories. Of these, 20 studies showed significantly high correlations, indicating that the subscales were not measuring independent traits. Since Ellis does not give specific references, it was impossible to check which of the 21 studies applied specifically to the MMPI.

A check of the literature, however, shows only two studies in which intercorrelations of MMPI subscales were reported. Only one of these studies had been published before Ellis' article appeared. Both studies, however, seem to support his contention that the subscales are correlated. But even a cursory inspection of the MMPI Manual [3] shows that no claim for freedom from intercorrelation has ever been made by the authors of the test. They point out that the correlation of the *Sc* scale with the *Pt* scale for normal cases is .84, but that it drops to .75 on abnormal cases. They state specifically that "The scale intercorrelations, which vary widely, will be difficult to interpret until more data are available on the dynamic inter-relationships of the different clinical syndromes" [3, p. 3].

It seems necessary here to make the elementary statistical observation that although two traits have a correlation of .70, this still means that more than 50 per cent of the variance is not taken into account by the correlation. Thus, while the subscales on the MMPI may have statistically significant intercorrelations, this does not mean that they do not have clinical usefulness either as separate subscales or as a group in profile analysis.

The second point that Ellis raises concerns the lack of agreement among the various inventories and their failure to correlate highly with the Rorschach and other projective techniques. Both the fact that there are such striking differences in the construction, reliability, and validity of the various personality inventories and the fact that the MMPI is an empirically developed key cast serious doubt upon the propriety of using other inventories as a check on its validity. Space does not permit a discussion of validity studies of the Rorschach and other projective techniques, but to say that their validity has not yet firmly been established experimentally seems to err on the side of charity.

The third objection raised by Ellis is that personality inventories are easily faked. Although he admits that the MMPI can partially compensate for this, we felt that a thorough examination of the studies which pertain to faking would be of value. Using the bibliography listed in the MMPI Atlas [4] for a guide, we were able to find nine relevant studies, eight of which showed that faking could be detected at a statistically significant level.

---

[1] Now at Michigan State College.

Ellis' next objection is perhaps the most serious. He states that personality inventories "usually do not give significant group discriminations when used with vocational, academic, socioeconomic, and disabled and ill groups. It was especially found that, in none of the areas in which they are commonly employed, do personality inventories consistently show significant group discriminations" [2, p. 48]. He supports this statement by an analysis of 499 studies of personality inventories. Since he again fails to give references to the particular studies he used, we decided to examine all

Table 1
Significant Discriminations Found in Research Studies
with the MMPI, 1940 through 1950

Discriminations tested*	Number of studies in which significant discriminations were found	Number of studies in which no significant discriminations were found	Total number of studies
Inventory scores vs. diagnostic examinations:			
Neuropsychiatric conditions	14	4	18
Psychosomatic conditions	1	0	1
Delinquent, criminal, and psychopathic behavior	5	1	6
Alcoholism	4	0	4
Predicting success in treatment	9	0	9
Total	33	5	38
Inventory scores vs. behavioral characteristics:			
Extent of social participation	4	0	4
Tendency to become a sexual deviant	1	0	1
Success at motor tasks	1	0	1
Miscellaneous behavioral characteristics	5	0	5
Total	11	0	11
Inventory scores vs. vocational and academic test and performance results:			
Tendency to have specific vocational interests	3	0	3
Membership in different vocational groups	2	2	4
Success in actual vocational performance	4	2	6
Success in academic achievement	3	0	3
Total	12	4	15
Inventory scores vs. group differences:			
Intellectual groups	3	0	3
Sex groups	2	0	2
Socioeconomic groups	3	0	3
Disabled and ill groups	2	0	2
Miscellaneous groups	5	0	5
Total	15	0	15
Grand Total	71	9	80

* This table is a duplicate of that presented by Ellis [2]. We have omitted those categories of his for which we could find no relevant studies on the MMPI.

the published studies available to us which were relevant to the categories he lists in his table to see how the MMPI would hold up under this type of examination. Since the Atlas includes most papers which make more than a passing reference to the MMPI from 1940 to the end of 1950, and since the studies Ellis lists are from the period 1946 to 1951, we felt that most of the pertinent studies would be in this list.

The data presented in Table 1 seem to show that Ellis is in serious error in his belief that the MMPI does not *consistently* show significant group discriminations and that, as Ellis puts it, different experimenters keep obtaining directly contradictory results when using the same or similar inventories.

Ellis' final criticism is, we think, the most provocative. He feels that when personality inventories are most effectively used, they tend to become equally as time-consuming as alternative psychodiagnostic procedures (such as interviewing and the use of projective techniques) which seem to be more clinically incisive and valuable. This complaint reminds us of Dallenbach's oft-quoted remark, "Ease and convenience are poor experimental guides" [1]. It scarcely seems necessary to add that they also make for poor clinical criteria.

This belief that the clinical psychologist can gain more pertinent, incisive, and "depth-centered personality" material from more intuitive techniques than from the MMPI calls for careful examination. Again space does not permit a lengthy discussion of the validity of these intuitive techniques. We would, however, like to mention briefly a recent study by Holtzman and Sells [5].

These authors have presented us with an excellent study of the efficacy of present-day clinical techniques. They administered a battery of tests to a group of aviation cadets. The protocols were then presented to some of the most outstanding clinical psychologists in the country. Some of the cadets had adjusted to the stress of flight training while others were eliminated from the program because they developed overt personality disturbances. The clinicians were asked to separate one group from the other on the basis of the protocols.

The results of this experiment were that no clinician or combination of clinicians using any single test or any combination of tests could predict above the chance level.[2] But the most disturbing finding of this study was the interclinician variability. To read that one group of clinicians in general described a cadet as having strong ego strength or a good heterosexual adjustment pattern, and then to read that other clinicians felt in general that the *same* cadet evidenced latent schizophrenic trends or latent homosexual tendencies causes us to lose a good deal of our faith in clinical intuition and in the tests employed in this study.

The results of the Holtzman-Sells experiment, when compared to the findings listed in Table 1, lead us to believe that Ellis might improve his clinical intuition by laying aside his projective techniques for a furtive glance at an MMPI profile.

Received April 6, 1953.

References

1. Dallenbach, K. M. The temperature spots and end-organs. *Amer. J. Psychol.*, 1927, **39**, 402-427.
2. Ellis, A. Recent research with personality inventories. *J. consult. Psychol.*, 1953, **17**, 45-49.
3. Hathaway, S. R., & McKinley, J. C. *Manual for the Minnesota Multiphasic Personality Inventory.* New York: Psychological Corp., 1943.
4. Hathaway, S. R., & Meehl, P. E. *An atlas for the clinical use of the Minnesota Multiphasic Personality Inventory.* Minneapolis: Univer. of Minnesota Press, 1951.
5. Holtzman, W. H., & Sells, S. B. Prediction of adjustment in the USAF flight training program: clinical assessment of selected test protocols. Paper read at Amer. Psychol. Assn., Washington, D. C., August, 1952.

---

[2] Except for one clinician whose wrong guesses were so predominant that they reached the 5 per cent level of significance in the wrong direction.

NEW BOOKS AND TESTS

Books

Buros, Oscar Krisen. (Ed.) *The fourth mental measurements yearbook.* Highland Park, N. J. (220 Montgomery St.): Gryphon Press, 1953. Pp. xxiv + 1163. $18.00.

This volume covers the period 1948 through 1951, and attempts to list all commercially available tests—educational, psychological, and vocational—published as separates in English-speaking countries during this time. The list includes the staggering total of 793 tests, 130 more than the *Third Yearbook.* Many more personality tests have been included, together with a number of tests which are available only as a part of restricted testing programs. There are 596 original test reviews by 308 reviewers, 53 excerpts from journal test reviews, and 4,417 references on the construction, validity, use, and limitations of specific tests. In addition, 429 books on measurements and closely related fields are listed, and extensive excerpts from reviews of these books are reprinted. This volume, as have been the preceding ones, is magnificently organized, indexed, and cross-indexed. The complete bibliographies for each test are invaluable. Some criticisms of previous volumes have been well heeded, as the listed instructions to the test reviewers, and their performance, indicate. The level of the test reviews is generally high, and for the most part judicious. A few reviews are possibly excessively bitter. Most possibly controversial tests have two or more reviewers, and a comparison of the reviews serves to control the problem of possible personal bias of the reviewer; study of the reviews of several different tests by one reviewer may also be very illuminating here. Any prospective test user can certainly learn quickly, authoritatively, and painlessly what tests there are for any purpose, and what he needs to know about any test, so far as the knowledge is available anywhere. Would-be test constructors are urged to read and reconsider—maybe their brain child should be aborted. But if it must be brought to term, what should and should not be done is amply exemplified here. Read and heed!—*A. R.*

Devereux, George. (Ed.) *Psychoanalysis and the occult.* New York: International Universities Press, 1953. Pp. xv + 432. $3.00.

This is an anthology of widely differing views on "telepathic" and "clairvoyant" phenomena by psychoanalysts and psychoanalytically oriented writers. For the most part, the essays are based on the authors' observations of correspondences between thoughts of patients and analyst, or between patient's (or analyst's) thought and external events. The positions range from heated disbelief with explanations in terms of chance, suggestion, projection, faint sensory clues, etc., to fiery belief in the reality of thought transmission with arguments for nerve induction, identification, etc. Protagonist and antagonist each ascribe "deep unconscious sources" to the other's opposition. Calls for experimental validation by skeptics are often ignored by believers, or sometimes scoffed at, as when Fodor cries out: "Since when has the unconscious any respect for scientific evidence?" (p. 331). Meanwhile, editor Devereux looks out over this bloody battlefield with cool and often brilliant detachment. His relative impartiality is evident in the inclusion, in text and bibliography, of writers who stress the two extremes as well as several middle positions. But, additionally, he writes two chapters of his own that are characterized by neither haughty skepticism nor uncontrolled acceptance. Thus, Devereux displays some of the self-discipline and relative lack of bias that Freud demanded of himself and followers, but that neither Freud nor many followers, as this book amply demonstrates, could quite muster.—*E. L. K.*

Fliess, Robert. *The revival of interest in the dream.* New York: International Universities Press, 1953. Pp. 164. $3.00.

Many years ago Freud complained that analysts lacked interest in the dream and had little to add to dream theory. In this small volume a close student of Freud presents abstracts of recent selected papers on the dream, classified according to clinical observations, applied dream interpretation, or additions to dream theory. He includes as well his own contribution to the meaning of the "spoken word" in the dream. In each instance, Fliess appends his own critical note to the published papers, frequently pointing out that the points raised by

---

Note.—The reviews were prepared by the Associate Editors, who may be identified by their initials.

the contemporary writer were handled or at least recognized earlier by Freud. Indeed, Fliess seems unduly constrained to demonstrate that little has been contributed to dream theory since Freud's writings on the topic. If he does not entirely succeed in this, at least he shows conclusively that among analysts interest in the dream has been "revived." This is a useful source book of recent dream studies and a provocative pro-Freudian critique of dream theory.—*A. M. G.*

Hammond, Kenneth R., & Allen, Jeremiah M., Jr. *Writing clinical reports.* New York: Prentice-Hall, 1953. Pp. xii + 235. $5.35.

Practicing clinicians, as well as clinical teachers and students, often find report-writing difficult and reports dissatisfying and ambiguous. This manual, written jointly by a psychologist and a member of an English department, provides some real assistance. The philosophy of report-writing, the problems of communication with sophisticated and unsophisticated readers, and the effective use of technical and nontechnical vocabulary are particularly well handled. The sections on style and organization seem less directly helpful, although examples are consistently drawn from the psychological field. Many of the points made, and all of the extensive illustrative case material, are the outgrowth of suggestions provided by staffs of many different institutions polled by the authors. This book will find its way into the psychological library of most clinics and hospitals; it deserves consideration also by university teachers of clinical psychology.—*A. M. G.*

Kogan, Leonard S., Hunt, J. McVicker, & Bartelme, Phyllis F. *A follow-up study of the results of social casework.* New York (192 Lexington Ave.): Family Service Association of America, 1953. Pp. 115. $2.50.

This is the detailed report of an intensive follow-up study of the adjustment of 38 families five years after social casework treatment. All accessible members of these families were located and interviewed by a psychologist unacquainted with the nature of the families' earlier problems. Practically all former clients were found to place a high value on the help they had received. Judged improvement during the period of casework appears to have been sustained after closing of the cases but was not related to judged improvement after closing. Psychologists should find especially valuable the section dealing with methodological and design problems of follow-up studies.—*E. L. K.*

Lacey, Oliver L. *Statistical methods in experimentation.* New York: Macmillan, 1953. Pp. xi + 249. $4.50.

This is a brief introductory text designed to introduce students to the use of statistics in experimentation. It is clearly written and well developed. Its use, however, will come at an elementary, introductory level. It is not designed for the sophisticated experimenter.—*W. A. H.*

Lindgren, Henry C. *Psychology of personal and social adjustment.* New York: American Book Co., 1953. Pp. ix + 481. $4.50.

A book about normal people and their social relationships, which gives an excellent coverage of all phases of human adjustment and mental health. While based on psychological findings, the book is so well written that previous training in psychology is not needed to hold the student's understanding and interest. Each chapter is followed by a concise summary of the important points, together with a list of source materials and suggested readings. Salient points are strengthened by excellent illustrations and cartoons. There is a complete author and subject index. Although it is planned as a textbook for college students, the book should prove interesting reading for anyone wishing to gain a better understanding of everyday problems in human living.—*B. M. L.*

Moreno, J. L. *Who shall survive?* (Rev. Ed.) Beacon, N. Y.: Beacon House, 1953. Pp. cxiv + 763. $10.00.

In his 114-page "preludes of the sociometric movement," Moreno presents a professional autobiography and concludes it with this apparently cogent insight: "There is no controversy about my ideas; they are universally accepted. I am the controversy." There is no statement as to how this large volume differs from the 1934 edition, to which he refers here as a "bible of human relations" (p. lxvi). Comparison of the two indicates that this is an expansion of the basic outline with slightly more discussion of sociodrama, group psychotherapy, and psychodrama, although much less than one might expect from the subtitle and extensive bibliographies devoted to them. More attention is given in this book than in its predecessor to the broader implications of sociometry and sociometric and spontaneity theory. The original book presented the theoretical bases for sociometry and the use of sociometric techniques with various groups. Despite the fact that the bibliography is greatly expanded (1,300 titles as compared to 41 in the first edition) there is no attempt to integrate systematically this wealth of material into the text proper. The book is an unabashed exposition of Moreno's ideas and is limited largely to a consideration of their influences. Certainly Dr. Moreno has made a great and widely recognized contribution to the study and effective use of the group. However, the reviewer feels that these important ideas could have been presented with greater effect in fewer pages and would have been enriched by integrating them with the outstanding work of others which has appeared since the date of the first edition.—*F. McK.*

Naumberg, Margaret. *Psychoneurotic art: its function in psychotherapy.* New York: Grune & Stratton, 1953. Pp. x + 148. $6.75.

Through the medium of an intensive case study, Miss Naumberg describes a psychoanalytically oriented method of therapy which relies heavily upon the symbolic character of the patient's spontaneous art products. The presentation is perceptive and reflects the broad experience of a skilled clinician. Her own discussion is buttressed by an analysis by Adolph Woltmann of the Rorschach responses of a single case in relation to the art work produced during three years of therapy, and by Piotrowski's comments on the Rorschach generally as it relates to the therapeutic use of graphic art. A critical review of the literature on art in clinical practice forms a useful appendix. It is obviously vital to the clinical culture for new techniques and methods to circulate freely. This book admirably illustrates the kind of procedural description that will allow other practitioners to adapt Miss Naumberg's approach to their own therapeutic armamentaria. When it is concerned with theory, however, it is at times fuzzy, at times doctrinaire, failing to provide any adequate underpinning of ideas for the practical techniques it advances. More concretely, readers may worry about the soundness of the interpretations derived from art products, just as they may wonder at the applicability of the method to more than a fraction of the neurotic population. Miss Naumberg's book represents a genuinely creative effort. It deserves wide reading and the serious, critical consideration by which private hunches may be transformed into useful public knowledge.—*E. J. S.*

Peters, R. S. (Ed.) *Brett's history of psychology.* New York: Macmillan, 1953. Pp. 742. $7.50.

In condensing G. S. Brett's three volumes into one (Vol. I, 1912; Vols. II and III, 1921), Peters has retained the original organization and even the writing of Brett. Except for the last of 15 chapters, editing was done by omitting entire sections and adding introductory statements to each chapter. Even in its condensed form, this is a formidable volume. Except for the new chapter, it is more of a history of thought than of psychology. It traces in great detail the developments in philosophy, medicine, and the physical sciences that led to psychological formulations from earliest recorded history to the work of William James. In bringing the work up to date, Peters has discarded the original chapter on "modern" psychology, and substituted one on twentieth-century trends. This summary of the history of psychology for the past fifty years constitutes a thumbnail sketch of some of the more common schools of contemporary psychology. This chapter does not have the sweep of Brett's treatment of earlier history, but the editor was frank to recommend other works for fuller treatment of the psychology of the past half century. The abridgement of Brett's *History* renders an important service to psychology by making it more easily available to students who, in their involvement with the technical aspects of modern psychology, may not be aware of the origins of current movements.—*M. K.*

Piaget, Jean. *The origins of intelligence in children.* New York: International Universities Press, 1953. Pp. xi + 419. $6.00.

This book, taken together with *La Construction du Reel Chez l'Enfant,* and *La Formation du Symbole Chez l'Enfant,* constitutes a series of works devoted to the beginnings of intelligence. Piaget considers with particular care the infantile behavior which precedes, but forms the basis for, "intelligent" behavior: those elementary sensorimotor adaptations commonly classified as reflexes and elementary habits. He is then able, by means of the concepts of circular reaction and the formation of secondary schemata, to account for the development from simple sensorimotor adjustment to the discovery and invention of new means of adaptation. His emphasis on the importance of perceptual behavior in this sequence, and on the significance of learning (or "use") in promoting intellectual development, will strike a familiar note for American psychologists. As usual, Piaget shows in this work the careful, ingenious observation of infants, and the detailed reports of even their most fleeting reactions, which make his developmental studies so stimulating. It should be noted that the translation is excellent, rendering an already compelling text into clear, appealing, familiar terms. —*A. M. G.*

Senn, Milton J. E. (Ed.) *Problems of infancy and childhood.* Transactions of the Sixth Conference. New York: Josiah Macy, Jr. Foundation, 1953. Pp. 160. $2.50.

This published account of the Sixth Macy Conference on Infancy and Childhood brings together three extended discussions of problems of infancy, along with records of the group interchange following each presentation. Escalona reports her observations of emotional development and mother-infant relationship in the first year of life. K. M. Wolf presents some of the cases of mothers and infants studied in connection with the Yale rooming-in project. Stewart describes the studies of excessive crying in infants which are under way at the University of Washington Child Health Center. All three papers are notable for their contributions both to the phenomenology of infant behavior and to the development of dynamic hypotheses to account for it. Papers and discussion alike retain their original informal, conference form—a characteristic which adds markedly to their appeal and readability—*A. M. G.*

Slavson, S. R. *Child psychotherapy.* New York: Columbia Univer. Press, 1952. Pp. xiii + 332. $4.50.

This book is much more than its title suggests. The consideration of therapy with children is preceded by two important sections on normal child development in terms of needs, and on psychopathology in childhood, in terms of early primary relationships and fixations. The dynamics of therapy are then presented as an outgrowth of these views of normal and pathological development. The author acknowledges his indebtedness to Freud and maintains throughout a predominantly psychoanalytic orientation. As in his other books, so here Slavson probably makes his most significant contribution in his careful breakdown of the therapeutic process, his attention to the role of parents and close relatives in affecting child behavior, and his emphasis on group factors in illness and treatment. The brief case illustrations which occur in the text often seem oversimplified by contrast to the theoretical exposition. The somewhat longer therapeutic record which concludes the book, on the other hand, adds considerably to the reader's understanding of the principles involved.—*A. M. G.*

Townsend, John C. *Introduction to experimental method.* New York: McGraw-Hill, 1953. Pp. ix + 220. $4.00.

This book is intended to appeal to three categories of readers: "First, the undergraduate student who is undergoing his first exposure to the rigors of the experimental method in psychology and the social sciences. Second, the student who, with an inadequate background in the application of the experimental method, finds himself faced with the necessity of 'doing a piece of research' to satisfy thesis requirements. Third, the social science worker who discovers that his job in industry, the clinic, the prison, etc., is one demanding the execution of research projects." The writer's presentation is deliberately superficial; therefore the book's content may not be sufficient to the interests or the requirements of his second and third groups of readers. The material could prove useful, however, in undergraduate instruction, either as supplementary reading for the elementary psychology course or as part of the assigned material for the undergraduate laboratory. The content of the book is concerned with experimentation as traditionally employed in psychology. Chapters 2 and 9, which have to do with causal sequences and methods of inference, are likely to contribute to clear thinking among undergraduates without the care and tedium which comparable material in a book on the logic of scientific method would require. Two other chapters, 7 and 8, which have to do with the control of the experiment and procedure for experimentation, will be useful to the casual undergraduate student of psychology. Other chapters have to do with such matters as variables, controls, and apparatus. The meaning of theory and the problem of theory building, which are important in experimental psychology, are for the most part either ignored or evaded by reference to other works. This kind of book could appeal to a wider audience if it were possible to prepare it in a manner more scholarly, thorough, and explicit. Apparently, however, the author considers his present superficial treatment best for his purposes.—*J. R. W.*

Books Received

Berdie, Ralph F., Layton, Wilbur L., & Hagenah, Theda. *Using tests in counseling.* Minneapolis: Student Counseling Bureau of Univer. of Minn., 1953. Pp. v + 86. Paper, $1.00.

Berkman, Tessie D. *Practice of social workers in psychiatric hospitals and clinics.* New York (1860 Broadway): Amer. Assoc. of Psychiatric Social Workers, 1953. Pp. ix + 158. Paper, $2.00.

Bosselman, Beulah Chamberlain. *The troubled mind.* New York: Ronald Press, 1953. Pp. v + 206. $3.50.

Burgess, Ernest W., & Wallin, Paul. *Engagement and marriage.* Philadelphia: Lippincott, 1953. Pp. xii + 819.

Fremont-Smith, Frank. (Ed.) *Health and human relations.* Report of Josiah Macy Jr. Foundation Conference, August 2-7, 1951. New York: Blakiston, 1953. Pp. xx + 192. $4.00.

Gorlow, Leon, Hoch, Erasmus L., & Telschow, Earl F. *The nature of nondirective group psychotherapy.* New York: Teachers Coll., Columbia Univer., Bureau of Publications, 1952. Pp. xi + 143. $3.25.

Hoerr, Normand L., & Osol, Arthur. *Illustrated pocket medical dictionary.* New York: Blakiston, 1952. Pp. xvi + 1005 + 24 figs. $3.75.

Jersild, Arthur T. *In search of self.* Teachers Coll., Columbia Univer., Bureau of Publications, 1952. Pp. xii + 145. $2.75.

Jersild, Arthur T., Helfant, Kenneth, & associates. *Education for self-understanding.* New York: Teachers Coll., Columbia Univer., Bureau of Publications, 1953. Pp. ix + 54. Paper, 85c.

Katz, Barney, & Lehner, George F. J. *Mental hygiene in modern living.* New York: Ronald Press, 1953. Pp. x + 544. $4.50.

Kubie, Susan H., & Landau, Gertrude. *Group work with the aged.* New York: International Universities Press, 1953. Pp. 214. $3.50.

Mettler, Fred A. (Ed.) *Psychosurgical problems.* New York: Blakiston, 1952. Pp. xii + 357. $7.00.

Piaget, Jean. *The child's conception of number.* New York: Humanities Press, 1952. Pp. ix + 248. $5.00.

Recktenwald, Lester Nicholas. *Guidance and counseling.* Washington, D. C.: Catholic Univer. of Amer. Press, 1953. Pp. xv + 192.

INDEX TO VOLUME 17

Subject Index

- ACE, and Wechsler-Bellevue, 63
- Achievement motivation, measured by picture interpretations, 289
- Adjustment, and human figure drawings, 421
  - and self concept, 39
  - and values, index of, validity, 135
  - index based on *Q* sort, 339
- Age, and perceptual learning, 222
  - and Wechsler-Bellevue reliabilities, 225
- Anxiety, effect on memory span, 265
  - manifest, changes with client-centered therapy, 443
  - Taylor Scale, forced-choice form, 447
- Autonomic reactions, and frustration, 267

- Bender-Gestalt test, scoring for directionality of lines, 213
- Body-cathexis, and self, 343
- Books received, 78, 156, 234, 320, 402, 468
- Book reviews, 76, 154, 232, 318, 399, 465

- Case histories, prognostic criteria in, 163
- Cattell Infant Intelligence Scale, and clinical judgment, 303
- Children's Apperception Test, variations in ambiguity, 67
- Client-centered therapy, and manifest anxiety changes, 443
  - concomitant MMPI changes, 334
  - for delayed speech, 122
  - *see also* nondirective therapy
- Clinical judgment, in measuring infant intelligence, 303
- Clinical report, utility, 297

- Deaf children, and WISC performance scale, 396
- Delinquency, follow-up study, 309
- Draw-a-Person test, sex sequence and MMPI M-F scale, 71

- Ego strength, scale predicting response to therapy, 327
- Emotionality, and Index of Adjustment and Values, 135
- Empathy and insight test, evaluation, 172
- Epilepsy, personality factors, 429
- Ethnocentrism, changes with psychotherapy, 235

- Face, sex differences and cultural preferences in response to, 126
- Factor analysis, parent behavior, 306
  - Rorschach, 261
  - schizophrenic symptoms, 455
- Fels Parent Behavior scales, factor analysis of, 306
- Frontal lobe operations, changes following, 92

- Group psychotherapy *see* Psychotherapy

- HTP, validity of traumatic indicator, 74
- Human figure drawings, and adjustment, 421

- Illusions, visual, and memory ability, 317
- Index of Adjustment and Values, and level of aspiration, 206
  - changes in scores related to emotionality, 135
- Insight and empathy test, evaluation, 172
- Intelligence, deficit in schizophrenics and organics, 389
  - infant, measurement, 303
  - relation to Rorschach scores, 21
  - social and general, and Rosenzweig P-F study, 293
- Interview, analysis of larger units, 184
  - analysis of pauses, 1
  - rating therapist's responses, 321
- Involutional melancholia, and Rorschach, diagnosis, 365
- I. P. A. T. Junior Personality Quiz, 436

- Junior Personality Quiz, construction of, 436

- Learning, perceptual, relation to age, 222
- Leucotomy, changes following, 92
- Level of aspiration, and Index of Adjustment and Values, 206
- Levy Movement Test, and Rorschach movement responses, 195

- Madeleine Thomas Completion Stories Test, description of, 139
- Memory ability, and shock therapy, 317
  - and visual illusions, 317
- Memory span, effect of anxiety, 265
- Minnesota Multiphasic Personality Inventory, and changes with psychotherapy, 235, 334
  - and Draw-a-Person Test, 71
  - and measuring adjustment, 39
  - and personality characteristics, 142
  - ego-strength scale predicting response to therapy, 327
  - evaluation, 462
  - *K* scale, note on, 314
  - psychopathic deviate scale, validity, 459
  - test-retest stability, 217

- Need achievement *see* Achievement motivation
- Neurocirculatory reactions, and frustration, 267
- Neurosis, diagnosis with Rorschach, 25, 110
- Nondirective play therapy, research in, 177
  - *see also* Client-centered therapy
- Nondirective therapy, used with shock therapy, 87
- Nonintellectual Intelligence Test, 242

- Organic psychosis, intellectual deficit, 389

- Parent behavior, factor analysis of, 306
- Personality inventories, evaluation of recent research, 45, 462

Play therapy, nondirective, research in, 177
Prefrontal lobotomy, and Wechsler-Bellevue scores, 229
Problem-solving, generalization of rigidity, 100
Projective techniques, ambiguity of stimuli, 283
  Children's Apperception Test, varying ambiguity in, 67
  use of sounds, 377
Psychosis, diagnosis by Rorschach, 110
  intellectual deficit, 389
  prognosis with case histories, 163
  schizophrenia and overinclusive thinking, 384
  and shock therapy, 79
  and group psychotherapy, 113
Psychotherapy, and the clinical report, 297
  analysis of pauses in interviews, 1
  for delayed speech, 122
  group, with psychotic patients, 113
  MMPI changes, 334
  predicting continuation with Rorschach, 16, 104
  predicting duration with Rorschach, 348
  predicting responses by ego-strength scale, 327
  predicting results with Rorschach, 8
  rating therapist's responses, 321
  techniques applied to diagnostic testing, 132
  test correlates, 235
  use of *Q* sorts in, 339

Q technique, and index of adjustment, 339

Rigidity, effects of shock therapy, 79
  in problem solving, 100
Rorschach, and effects of stress, 247, 355
  and "eye" content response, 189
  and Levy Movement Test, 195
  and process-reactive schizophrenia, 157
  and schizophrenia, cross-validation, 371
  and self acceptance, 36
  changes with psychotherapy, 235
  diagnosing involutional melancholia, 365
  diagnostic signs, 110
  discriminating neurotics and schizophrenics, 25
  effect of order of card presentation, 32
  examiner's influence, 425
  factor analysis, 261
  human movement responses, classification, 254
  predicting continuation in therapy, 16, 104
  predicting duration of therapy, 348
  predicting prognosis in paranoid schizophrenics, 199
  predicting results in therapy, 8
  relation of free association and inquiry, 359
  relation of scores to intelligence and verbal fluency, 21
Rosenzweig Picture-Frustration Study, and social and general intelligence, 293
Rosenzweig triadic hypothesis, and seizure states, 429

Schizophrenia, diagnosis by Rorschach, 25
  distorted thought processes, 50
  effects of sodium amytal, 54
  factor analysis, 455
  intellectual deficit, 389
  overinclusive thinking, 384
  process and reactive, 157
  prognosis predicted by Rorschach, 199
  Rorschach patterns, 371
Seizure states, personality factors, 429
Self acceptance, and Rorschach scores, 36
Self, and body-cathexis, 343
Self concept, relation to inferred self and adjustment, 39
Shock therapy, and memory ability, 317
  effect on rigidity, 79
  relapses during, 87
Socioeconomic status, and WISC, 58
Sodium amytal, effect on schizophrenia, 54
Stanford-Binet, IQ changes in, 153
  and WISC, 147
Stress, and neurocirculatory reactions, 267
  effect on Rorschach scores, 247, 355

Taylor Manifest Anxiety Scale, forced-choice scale, 447
Testing, communication and rapport in, 132
Tests, ACE, 63
  Bender-Gestalt, 213
  Draw-a-Person, 71
  HTP, 74
  Insight and empathy, 172
  Junior Personality Quiz, 436
  Levy Movement, 195
  Madeleine Thomas Completion Stories, 139
  Nonintellectual Intelligence, 242
  Stanford-Binet, 147, 153
  Wechsler-Bellevue, 63, 389, 403, 411, 419
  WISC, 58, 147
  reviews, 78, 156, 233, 319, 401
Therapy *see* Psychotherapy
Thematic Apperception Test, ambiguity and personality factors, 283
  Negro and white responses, 373
  women's responses, 269
Topectomy, changes following, 92

Verbal fluency, relation to Rorschach scores, 21

Wechsler-Bellevue, and ACE, 63
  and schizophrenia and organic psychosis, 389
  changes after prefrontal lobotomy, 229
  changes with psychotherapy, 235
  ranking subtests for diagnosis, 403
  reliability, age differences, 225
  sex differences in superior young adults, 411
  variability of full-scale IQ and verbal and performance IQ, 419
Wechsler Intelligence Scale for children, and socioeconomic status, 58
  and Stanford-Binet, 147
  performance scale, deaf children, 396
Women, TAT responses, 269

Author Index

The names of authors of original articles are printed in *italics*; those of authors of books and tests reviewed in Roman.

Abramson, H. A., 232
Abt, L. E., 318
Allen, J. M., Jr., 466
Allen, R. M., 195
Anderson, Helen J., 459
Anderson, Rose G., 320
Armitage, S. G., 50
Ashby, W. R., 76
Auld, F., Jr., 104

Baker, H. J., 318
Barker, R. G., 399
Baumgarten, F., 76
Barrell, R. P., 254
Barron, F., 235, 327
Bartelme, Phyllis F., 466
Bayard, J., 163
Beier, E. G., 126
Bellak, L., 156
Bellak, Sonja S., 156
Beloff, H., 436
Bennett, G. K., 78, 401
Berger, D., 355
Berkowitz, M., 110
Bijou, S. W., 283
Bills, R. E., 36, 135, 206
Blum, G. S., 318
Blum, Lucille H., 233
Botwinick, J., 225
Bowers, H., 154
Brower, D., 318
Bruce, M. M., 401
Burlingham, Dorothy, 232
Buros, O. K., 465

Cabot, H., 318
Calvin, A. D., 39, 462
Canter, F. M., 429
Carlson, V. R., 247
Cattell, R. B., 436
Child, I. L., 401
Cole, D., 132
Cole, Mary E., 163
Collier, R. M., 321
Cook, D. M., 267
Cowen, E. L., 100
Crown, S., 92

Dailey, C. A., 297
Daston, P. G., 50
Davidson, Audrey, 232
Davidson, H. A., 154
Devereux, G., 465
Dupont, H. J., 122
Dymond, Rosalind, F., 339

Ellis, A., 45
Eissler, Ruth S., 154
Epstein, S., 384
Eron, L. D., 104, 269
Estes, Betsy Worth, 58
Eysenck, H. J., 154

Fay, Judith, 232
Feifel, H., 113
Feldman, Dorothy A., 163
Fiedler, Miriam F., 155
Fieldsteel, Nina D., 233
Fleiss, R., 465
Frosch, J., 76

Galanter, E. H., 74
Gallagher, J. J., 303, 334, 443
Garrett, H. E., 232
Garrett, J. F., 232
Gelink, Marjorie, 401
Gibby, R. G., 348, 359, 425
Gilbert, Jeanne G., 155
Goldhamer, H., 399
Gonick, Mollie R., 399
Gough, H. G., 242
Graham, E. E., 396
Granick, S., 71
Grassi, J. R., 319
Grauer, D., 199
Gross, Margaret V., 319
Guertin, W. H., 455

Hagan, Helen R., 400
Hammond, K. R., 8
Hammond, K. R., 466
Hamrin, S. A., 319
Hanes, B., 222
Harrower, Molly, 233
Heathers, L. B., 63
Heineman, C. E., 447
Hess, Judith, 100
Hiler, E. W., 348
Hilgard, E. R., 155
Hirsh, I. J., 77
Holland, G. A., 147
Holtzman, W. H., 39
Hovey, H. B., 142
Hoyles, J. A., 319
Hunt, J. McV., 466
Husni, May, 377

Izard, C. E., 126

Jastak, J., 403
Jay, Edith S., 156
Jenkins, R. L., 306
Jourard, S. M., 343

Kahl, J. A., 318
Kantor, R. E., 157
Karlin, L., 293
Karpf, F. B., 399
Kaye, D., 265
Kenny, D. T., 283
Kessler, H. H., 399
King, G. F., 50
Kirschner, P., 265
Kitzinger, Helen, 229
Kluckhohn, C., 233

Kooler, F. J., 365
Kogan, L. S., 466
Kornitzer, Margaret, 77
Kotkov, B., 16
Krieg, W. J. L., 155
Kubie, L. S., 155
Kuhlen, R. G., 77
Kuhlmann, F., 320

Lacey, O. L., 466
Landsman, T., 122
Lawrence, J. F., 261
Lazarus, R. S., 156
Lazarus, R. S., 247
Lebo, D., 177
Levine, J., 110
Levine, M., 74
Lindgren, H. C., 172
Lindgren, H. C., 466
Lindquist, E. F., 399
Lonstein, M., 371
Lorr, M., 306
Lotsof, E. J., 21
Lynn, D. B., 67

McConnell, J., 462
McNemar, Q., 153
Machover, S., 459
Mainord, W. A., 54
Mandler, G., 265
Maradie, L. J., 32
Markwell, E. D., Jr., 229
Marshall, A. W., 399
Maslow, A. H., 234
Meadow, A., 16
Merrill, R. M., 63
Meyerson, L., 399
Miller, D. R., 348, 425
Mills, E. S., 139
Moreno, J. L., 466
Morgan, H. H., 289
Mowrer, O. H., 400
Moustakas, C. E., 233
Murray, H. A., 233
Mussen, P. H., 373
Muthard, J. E., 184

Naumberg, Margaret, 467
Norman, R. D., 411

Overholser, W., 400

Pascal, G. R., 163
Patterson, C. H., 319
Peek, R. M., 213
Peters, R. S., 467
Piaget, J., 467
Podolsky, E., 400
Poole, R. C., 195
Pugh, R. W., 87
Pullen, M. S., 79
Pumpian-Mindlin, E., 155

Quay, H., 314
Quiroz, F., 389

Rappaport, S. R., 389
Ray, C. D., 195
Reid, J. H., 400
Reik, T., 155
Rieman, G. W., 25
Robinson, Jacqueline, 172
Rogers, L. S., 8
Rosen, A., 217
Rubin, H., 371

Sarason, S. B., 233
Schneider, D. M., 233
Schwartz, A. D., 113
Schwartz, M. M., 293
Seashore, H. G., 78
Secord, P. F., 343
Senn, M. J. E., 467
Shaffer, G. W., 156
Shapiro, Esther, 396
Skinner, B. F., 233
Slavson, S. R., 468
Smith, L. J., 71
Smock, C. D., 126
Stagner, R., 79
Standlee, L. S., 317
Stanley, J. C., 419
Stiel, Agnes, 365
Stotsky, B. A., 348, 359
Strother, G. B., 267
Sullivan, H. S., 401
Sweetland, A., 314
Swensen, C. H., 163

Thompson, G. G., 77
Toman, W., 1
Tougas, R. R., 126
Townsend, J. C., 468
Trenaman, J., 319

U. S. Federal Security Agency, 76

Valentine, M., 122

Walker, E. L., 425
Wallner, J. M., 157
Wattenberg, W. W., 309
Wechsler, D., 156
Weinland, J. D., 319
Weisskopf-Joelson, Edith A., 67
Wertheimer, M., 189
Wesman, A. G., 78
Wheeler, W. M., 229
Whitmyre, J. W., 421
Whiting, J. W. M., 401
Wiener, M., 100
Williams, H. L., 261
Wilmer, H. A., 377
Winder, C. L., 157

Wolfle, D., 77
Wood, H. B., 401
Wright, Beatrice A., 399

Zilaitis, V., 455